cited in the European Search Report of EP-97906876.4 Your Ref.: 090101 012 EP1

WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6:

C07H 21/04, C12N 1/20, 1/14, 5/00, 9/38, 9/42, C08B 30/04

(11) International Publication Number:

WO 98/24799

(43) International Publication Date:

11 June 1998 (11.06.98)

(21) International Application Number:

PCT/US97/22623

A1

(22) International Filing Date:

8 December 1997 (08.12.97)

(30) Priority Data:

60/056,916 Not furnished 6 December 1996 (06.12.96)

US US 10 October 1997 (10.10.97)

(71) Applicant (for all designated States except US): DIVERSA CORPORATION [US/US]; 10665 Sorrento Valley Road, San Diego, CA 92121 (US).

(72) Inventors; and

- (75) Inventors/Applicants (for US only): BYLINA, Edward, J. [US/US]; Apartment A-1, West Court, Andalusia, PA 19020 (US). SWANSON, Ronald, V. [US/US]; Apartment A, 309 No. Lemon Street, Media, PA 19063 (US). MATHUR, Eric, J. [US/US]; 2654 Galicia Way, Carlsbad, CA 92009 (US). LAM, David, E. [US/US]; 1518 West 249th Street, Harbor City, CA 90710 (US).
- (74) Agent: HAILE, Lisa, A.; Fish & Richardson P.C., Suite 1400, 4225 Executive Square, La Jolla, CA 92037 (US).

(81) Designated States: AU, CA, IL, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: GLYCOSIDASE ENZYMES

(57) Abstract

Thermostable glycosidase enzymes derived from various Thermococcus, Staphylothermus and Pyrococcus organisms is disclosed. The enzymes are produced from native or recombinant host cells and can be utilized in the food processing industry, pharmaceutical industry and in the textile industry, detergent industry and in the baking industry.

cred in the hardware Spatest Report of the Your Press.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Stovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	
AU	Australia	GA	Gabon	LV	Latvia	SZ	Senegal
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Swaziland
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Chad
BB	Barbados	GH	Ghana	MG	Madagascar		Togo
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TJ .	Tajikistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TM	Turkmenistan
BG	Bulgaria	HU	Hungary	ML	Mali	TR	Turkey
BJ	Benin	IE	Ireland	MN	Mongolia	TT	Trinidad and Tobago
BR	Brazil	IL	İsrael	MR	Mauritania	UA	Ukraine
BY	Belarus	IS	Iceland	MW	Malawi	UG	Uganda
CA	Canada	IT	Italy	MX	Mexico	US	United States of Americ
CF	Central African Republic	JP	Japan	· NE	Niger	UZ	Uzbekistan
CG	Congo	KE	Kenya	NL	Netherlands	VN	Viet Nam
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	YU	Yugoslavia
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand	zw	Zimbabwe
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic -	LC	Saint Lucia	RU			
DE	Germany	Li	Liechtenstein	SD	Russian Federation Sudan		
DK	Denmark	LK	Sri Lanka	SE			•
EE	Estonia	LR	Liberia	SG	Sweden Singapore		

GLYCOSIDASE ENZYMES

BACKGROUND OF THE INVENTION

1. Field of the Inventions

This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, the use of such polynucleotides and polypeptides, as well as the production and isolation of such polynucleotides and polypeptides. More particularly, the polynucleotides and polypeptides of the present invention has been putatively identified as glucosidases, α -galactosidases, β -galactosidases, β -mannosidases, β -mannases, endoglucanases, and pullalanases.

2. Description of Related Art

The glycosidic bond of β -galactosides can be cleaved by different classes of enzymes: (i) phospho-β-galactosidases (EC3.2.1.85) are specific for a phosphorylated substrate generated via phosphoenolpyruvate phosphotransferase system (PTS)-dependent uptake; (ii) typical β-galactosidases (EC 3.2.1.23), represented by the Escherichia coli LacZ enzyme, which are relatively specific for β-galactosides; and (iii) β-glucosidases (EC 3.2.1.21) such as the enzymes of Agrobacterium faecalis, Clostridium thermocellum, Pyrococcus furiosus or Sulfolobus solfataricus (Day, A.G. and Withers, S.G., (1986) Purification and characterization of a β-glucosidase from Alcaligenes faecalis. Can. J. Biochem. Cell. Biol. 64, 914-922; Kengen, S.W.M., et al. (1993) Eur. J. Biochem., 213, 305-312; Ait, N., Cruezet, N. and Cattaneo, J. (1982) Properties of β-glucosidase purified from Clostridium thermocellum. J. Gen. Microbiol. 128, 569-577; Grogan, D.W. (1991) Evidence that β -galactosidase of Sulfolobus solfataricus is only one of several activities of a thermostable β-D-glycodiase. Appl. Environ. Microbiol. 57, 1644-1649). Members of the latter group, although highly specific with respect to the β-anomeric configuration of the glycosidic linkage, often display a rather relaxed substrate specificity and hydrolyze βglucosides as well as β -fucosides and β -galactosides.

10

15

20

Generally. α-galactosidases are enzymes that catalyze the hydrolysis of galactose groups on a polysaccharide backbone or hydrolyze the cleavage of di- or oligosaccharides comprising galactose.

Generally, \(\beta\)-mannanases are enzymes that catalyze the hydrolysis of mannose groups internally on a polysaccharide backbone or hydrolyze the cleavage of di- or oligosaccaharides comprising mannose groups. \(\beta\)-mannosidases hydrolyze non-reducing, terminal mannose residues on a mannose-containing polysaccharide and the cleavage of di- or oligosaccaharides comprising mannose groups.

Guar gum is a branched galactomannan polysaccharide composed of β -1,4 linked mannose backbone with α -1,6 linked galactose side chains. The enzymes required for the degradation of guar are β -mannanase, β -mannosidase and α -galactosidase. β -mannanase hydrolyses the mannose backbone internally and β -mannosidase hydrolyses non-reducing, terminal mannose residues. α -galactosidase hydrolyses α -linked galactose groups.

Galactomannan polysaccharides and the enzymes that degrade them have a variety of applications. Guar is commonly used as a thickening agent in food and is utilized in hydraulic fracturing in oil and gas recovery. Consequently, galactomannanases are industrially relevant for the degradation and modification of guar. Furthermore, a need exists for thermostable galactomannases that are active in extreme conditions associated with drilling and well stimulation.

There are other applications for these enzymes in various industries, such as in the beet sugar industry. 20-30% of the domestic U.S. sucrose consumption is sucrose from sugar beets. Raw beet sugar can contain a small amount of raffinose when the sugar beets are stored before processing and rotting begins to set in. Raffinose inhibits the crystallization of sucrose and also constitutes a hidden quantity of sucrose. Thus, there is ment to eliminating raffinose from raw beet sugar. α -Galactosidase has also been used as a digestive aid to break down raffinose, stachyose, and verbascose in such foods as beans and other gassy foods.

 β -galactosidases which are active and stable at high temperatures appear to be superior enzymes for the production of lactose-free dietary milk products (Chaplin, M.F.

10

15

20

and Bucke, C. (1990) In: Enzyme Technology, pp. 159-160, Cambridge University Press, Cambridge, UK). Also, several studies have demonstrated the applicability of β-galactosidases to the enzymatic synthesis of oligosaccharides via transglycosylation reactions (Nilsson, K.G.I. (1988) Enzymatic synthesis of oligosaccharides. Trends Biotechnol. 6, 156-264; Cote, G.L. and Tao, B.Y. (1990) Oligosaccharide synthesis by enzymatic transglycosylation. Glycoconjugate J. 7, 145-162). Despite the commercial potential, only a few β-galactosidases of thermophiles have been characterized so far. Two genes reported are β-galactoside-cleaving enzymes of the hyperthermophilic bacterium *Thermotoga maritima*, one of the most thermophilic organotrophic eubacteria described to date (Huber, R., Langworthy, T.A., König, H., Thomm, M., Woese, C.R., Sleytr, U.B. and Stetter, K.O. (1986) *T. martima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90 °C, Arch. Microbiol. 144, 324-333) one of the most thermophilic organotrophic eubacteria described to date. The gene products have been identified as a β-galactosidase and a β-glucosidase.

Pullulanase is well known as a debranching enzyme of pullulan and starch. The enzyme hydrolyzes α -1,6-glucosidic linkages on these polymers. Starch degradation for the production or sweeteners (glucose or maltose) is a very important industrial application of this enzyme. The degradation of starch is developed in two stages. The first stage involves the liquefaction of the substrate with α -amylase, and the second stage, or saccharification stage, is performed by β -amylase with pullalanase added as a debranching enzyme, to obtain better yields.

Endoglucanases can be used in a variety of industrial applications. For instance, the endoglucanases of the present invention can hydrolyze the internal \(\beta -1,4-\text{glycosidic} \) bonds in cellulose, which may be used for the conversion of plant biomass into fuels and chemicals. Endoglucanases also have applications in detergent formulations, the textile industry, in animal feed, in waste treatment, and in the fruit juice and brewing industry for the clarification and extraction of juices.

10

15

20

5

10

15

20

25

Brief Description of the Drawings

The following drawings are illustrative of embodiments of the invention and are not meant to limit the scope of the invention as encompassed by the claims.

Figures 1a-b are the full-length DNA and corresponding deduced amino acid sequence of M11TL of the present invention. Sequencing was performed using a 378 automated DNA sequencer for all sequences of the present invention (Applied Biosystems, Inc.).

Figure 2 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of OC1/4V-33B/G.

Figure 3 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of F1-12G.

Figures 4a-b are the full-length DNA and corresponding deduced amino acid sequence of 9N2-31B/G.

Figures 5a-b are the full-length DNA and corresponding deduced amino acid sequence of MSB8-6G.

Figure 6 is the full-length DNA and corresponding deduced amino acid sequence of AEDII12RA-18B/G.

Figures 7a-b are the full-length DNA and corresponding deduced amino acid sequence of GC74-22G.

Figures 8a-b are the full-length DNA and corresponding deduced amino acid sequence of VC1-7G1.

Figures 9a-c are the full-length DNA and corresponding deduced amino acid sequence of 37GP1.

Figures 10a-c are the full-length DNA and corresponding deduced amino acid sequence of 6GC2.

Figures 11a-d are the full-length DNA and corresponding deduced amino acid sequence of 6GP2.

Figures 12a-c are the full-length DNA and corresponding deduced amino acid sequence of 63GB1.

Figures 13a-b are the full-length DNA and corresponding deduced amino acid sequence of OC1/4V.

Figures 14a-e are the full-length DNA and corresponding deduced amino acid sequence of 6GP3.

Figures 15a-d are the full-length DNA and corresponding deduced amino acid sequence of *Thermotoga maritima* MSB8-6GP2.

Figures 16a-c are the full-length DNA and corresponding deduced amino acid sequence of *Thermotoga maritima* MSB8-6GB4.

Figures 17a-d are the full-length DNA and corresponding deduced amino acid sequence of *Banki gouldi* 37GP4.

Figures 18a-b are the full-length DNA and corresponding deduced amino acid sequence of *Pyrococcus furiosus* VC1-7EG1.

SUMMARY OF THE INVENTION

In a preferred embodiment of the present invention, there are provided isolated nucleic acids (polynucleotides) which encode mature enzymes having the deduced amino acid sequences of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64).

In another embodiment, the invention provides a method for producing a polypeptide including culturing host cells containing the polynucleotide of Figures 1-18 and expressing from the host cell a polypeptide encoded by the polynucleotide and isolating the polypeptide.

In another embodiment, the invention provides an enzyme selected from the group consisting of an enzyme having an amino acid sequence set forth in SEQ ID NOS: 15-28 or 61-64 and an enzyme which has at least 30 consecutive amino acid residue as an enzyme having an amino acid sequence set forth in SEQ ID NOS: 15-28 or 61-64.

In yet another embodiment, the invention provides a method for generating glucose from soluble cell oligosaccharides which includes contacting a sample containing oligosaccharides with an effective amount of an enzyme selected from the group of

5.

10

15

20

enzymes having the amino acid sequence set forth in SEQ ID NOS: 15-28, 61-63 and 64 such that glucose is produced

The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

Definitions

"Monosaccharide", as used herein, refers to a single polyhydroxy aldehyde or ketone unit.

"Oligosaccharide", as used herein, consist of short chains of monosaccharide units joined together by covalent bonds. Of these, the most abundant are the disaccharides, which have two monosaccharide units.

"Polysaccharide", as used herein, consists of long chains having many monosaccharide units.

The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

A coding sequence is "operably linked to" another coding sequence when RNA polymerase will transcribe the two coding sequences into a single mRNA, which is then translated into a single polypeptide having amino acids derived from both coding sequences. The coding sequences need not be contiguous to one another so long as the expressed sequences ultimately process to produce the desired protein.

"Recombinant" enzymes refer to enzymes produced by recombinant DNA techniques; *i.e.*, produced from cells transformed by an exogenous DNA construct encoding the desired enzyme. "Synthetic" enzymes are those prepared by chemical synthesis.

A DNA "coding sequence of" or a "nucleotide sequence encoding" a particular enzyme, is a DNA sequence which is transcribed and translated into an enzyme when placed under the control of appropriate regulatory sequences.

10

15

20

Detailed Description of the Invention

The polynucleotides and polypeptides of the present invention have been identified as glucosidases. α -galactosidases. β -galactosidases, β -mannosidases, β -mannanases, endoglucanases, and pullalanases as a result of their enzymatic activity.

In accordance with one aspect of the present invention, there are provided novel enzymes, as well as active fragments, analogs and derivatives thereof.

In accordance with another aspect of the present invention, there are provided isolated nucleic acid molecules encoding the enzymes of the present invention including mRNAs, cDNAs, genomic DNAs as well as active analogs and fragments of such enzymes.

In accordance with yet a further aspect of the present invention, there is provided a process for producing such polypeptides by recombinant techniques comprising culturing recombinant prokaryotic and/or eukaryotic host cells, containing a nucleic acid sequence of the present invention, under conditions promoting expression of said enzymes and subsequent recovery of said enzymes.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotides encoding such enzymes for hydrolyzing lactose to galactose and glucose for use in the food processing industry, the pharmaceutical industry, for example, to treat intolerance to lactose, as a diagnostic reporter molecule, in com wet milling, in the fruit juice industry, in baking, in the textile industry and in the detergent industry.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes for hydrolyzing guar gum (a galactomannan polysaccharide) to remove non-reducing terminal mannose residues. Further polysaccharides such as galactomannan and the enzymes according to the invention that degrade them have a variety of applications. Guar gum is commonly used as a thickening agent in food and also is utilized in hydraulic fracturing in oil and gas recovery. Consequently, mannanases are industrially relevant for the degradation and modification of guar gums. Furthermore, a need exists for thermostable mannases that are active in extreme conditions associated with drilling and well stimulation.

10

15

20

In accordance with yet a further aspect of the present invention, there are also provided nucleic acid probes comprising nucleic acid molecules of sufficient length to specifically hybridize to a nucleic acid sequence of the present invention.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotides encoding such enzymes, for *in vitro* purposes related to scientific research. for example, to generate probes for identifying similar sequences which might encode similar enzymes from other organisms by using certain regions, *i.e.*, conserved sequence regions, of the nucleotide sequence.

These and other aspects of the present invention should be apparent to those skilled in the art from the teachings herein.

The polynucleotides of this invention were originally recovered from genomic gene libraries derived from the following organisms:

M11TL is a new species of *Desulfurococcus* isolated from Diamond Pool in Yellowstone National Park. The organism grows optimally at 85-88°C, pH 7.0 in a low salt medium containing yeast extract, peptone, and gelatin as substrates with a N₂/CO₂ gas phase.

OC1/4V is from the genus *Thermotoga*. The organism was isolated from Yellowstone National Park. It grows optimally at 75 °C in a low salt medium with cellulose as a substrate and N_2 in gas phase.

Pyrococcus furiosus VC1 and (7EG1) is from the genus Pyrococcus. VC1 was isolated from Vulcano, Italy. It grows optimally at 100°C in a high salt medium (marine) containing elemental sulfur, yeast extract, peptone and starch as substrates and N₂ in gas phase.

Staphylothermus marinus F1 is a from the genus Staphylothermus. F1 was isolated from Vulcano, Italy. It grows optimally at 85°C, pH 6.5 in high salt medium (marine) containing elemental sulfur and yeast extract as substrates and N₂ in gas phase.

10

15

20

Thermococcus 9N-2 is from the genus Thermococcus 9N-2 was isolated from diffuse vent fluid in the East Pacific Rise. It is a strict anaerobe that grows optimally at 87°C.

Thermotoga maritima MSB8 and MSB8 (Clone # 6GP2 and 6GB4) is from the genus Thermotogo, and was isolated from Vulcano, Italy. MSB8 grows optimally at 85°C, pH 6.5 in a high salt medium (marine) containing starch and yeast extract as substrates and N₂ in gas phase.

Thermococcus alcaliphilus AEDII12RA is from the genus Thermococcus. AEDII12RA grows optimally at 85°C, pH 9.5 in a high salt medium (marine) containing polysulfides and yeast extract as substrates and N_2 in gas phase.

Thermococcus chitonophagus GC74 is from the genus Thermococcus. GC74 grows optimally at 85°C, pH 6.0 in a high salt medium (marine) containing chitin, meat extract, elemental sulfur and yeast extract as substrates and N₂ in gas phase. AEPII 1a grows optimally at 85°C at pH 6.5 in marine medium under anaerobic conditions. It has many substrates. Bankia gouldi is from the genus Bankia.

Accordingly, the polynucleotides and enzymes encoded thereby are identified by the organism from which they were isolated, and are sometimes hereinafter referred to as "M11TL" (Figure 1 and SEQ ID NOS:1 and 15), "OC1/4V-33B/G" (Figure 2 and SEQ ID NOS:2 and 16), "F1-12G" (Figure 3 and SEQ ID NOS:3 and 17), "9N2-31B/G" (Figure 4 and SEQ ID NOS:4 and 18), "MSB8" (Figure 5 and SEQ ID NOS:5 and 19), "AEDII12RA-18B/G" (Figure 6 and SEQ ID NOS:6 and 20), "GC74-22G" (Figure 7 and SEQ ID NOS:7 and 21), "VC1-7G1" (Figure 8 and SEQ ID NOS:8 and 22), "37GP1" (Figure 9 and SEQ ID NOS: 9 and 23), "6GC2" (Figure 10 and SEQ ID NOS: 10 and 24), "6GP2" (Figure 11 and SEQ ID NOS:11 and 25), "AEPII 1a" (Figure 12 and SEQ ID NOS:12 and 26), "OC1/4V" (Figure 13 and SEQ ID NOS:13 and 27), and "6GP3" (Figure 14 and SEQ ID NOS:28), "MSB8-6GP2" (Figure 15 and SEQ ID NOS:57 and 61), "MSB8-6GB4"(Figure 16 and SEQ ID NOS:58 and 62),"VC1-7EG1"(Figure 17 and SEQ ID NOS:59 and 63), and 37GP4 (Figure 18 and SEQ ID NOS:60 and 64).

5

10

15

20

The polynucleotides and polypeptides of the present invention show identity at the nucleotide and protein level to known genes and proteins encoded thereby as shown in Table 1.

Table 1

	T		
	Gene/Protein with	Protein	Nucleic Acid
Cl			
Clone	Closest Homology	Identity	Identity
M11TL-29G	Sulfolobus sulfataricus	51%	55%
	DSM 1616/P1, β-		
	galactosidase		
OC1/4V-33B/G	Caldocellum	52%	57%
	saccharolyticum, β-		
	glucosidase		
Staphylothermus	Bacillus polymyxa, β-	36%	48%
marinus F1-12G	galactosidase		
Thermococcus 9N2-	Sulfolobus sulfataricus	51%	50%
31B/G	ATCC 49255/MT4, β-		
	galactosidase	,	
Thermotoga maritima	Clostridium thermocellum	45%	53%
MSB8-6G	bglB		
Thermococcus	Bacillus polymyxa, β-	34%	48%
AEDII12RA-18B/G	galactosidase		
Thermococcus	Sulfolobus sulfataricus	46%	54%
chitonophagus GC74-	ATCC 49255/MT4, β-		
22G	galactosidase		

10

Pyrococcus furiosus VC1-7G1	Sulfolobus sulfataricus/MT-4 β- galactosidase	46.4%	52.5%
Thermotoga maritima α-galactosidase (6GC2)	Pediococcus pentosaceaus α-galactosidase	49%	29%
Thermotoga maritima B-mannanase (6GP2)	Aspergillus aculeatus mannanase	56%	37%
AEPII 1a ß- mannosidase (63GB1)	Sulfolobus solfactaricus ß- galactosidase	78%	56%
OC1/4V endoglucanase (33GP1)	Clostridium thermocellum endo-1,4-ß-endoglucanase	65%	43%
Thermotoga maritima pullalanase (6GP3)	Caldocellum saccharolyticum α- destrom 6 glucanohydralase	72	53
Bankia gouldi mix Endoglucanase (37GP1)	None available		

The polynucleotides and enzymes of the present invention show homology to each other as shown in Table 2.

10

Table 2

Clone	Gene/Protein with Closest Homology	Protein Identity	Nucleic Acid Identity
Staphylothermus marinus F1-12G	Thermococcus AEDII12RA-18B/G, β- galactosidase, glucosidase	55%	57%
Thermococcus 9N2- 31B/G	Thermococcus chitonophagus GC74- 22G-glucosidase`	74%	66%
Pyrococcus furiosus VC1-7G1	Pyrococcus furiosus VC1- 7B/G β-galactosidase	46.4%	54%

All the clones identified in Tables 1 and 2 encode polypeptides which have α -glycosidase or β -glycosidase activity.

This invention, in addition to the isolated nucleic acid molecules encoding the enzymes of the present invention, also provide substantially similar sequences. Isolated nucleic acid sequences are substantially similar if: (i) they are capable of hybridizing under conditions hereinafter described, to the polynucleotides of SEQ ID NOS: 1-14 and 57-60; (ii) or they encode DNA sequences which are degenerate to the polynucleotides of SEQ ID NOS: 1-14 and 57-60. Degenerate DNA sequences encode the amino acid sequences of SEQ ID NOS:15-28 and 61-64, but have variations in the nucleotide coding sequences. As used herein, substantially similar refers to the sequences having similar identity to the sequences of the instant invention. The nucleotide sequences that are substantially the same can be identified by hybridization or by sequence comparison. Enzyme sequences that are substantially the same can be identified by one or more of the following: proteolytic digestion, gel electrophoresis and/or microsequencing.

One means for isolating the nucleic acid molecules encoding the enzymes of the present invention is to probe a gene library with a natural or artificially designed probe using art recognized procedures (see, for example: Current Protocols in Molecular Biology,

15

20

Ausubel F.M. et al. (EDS.) Green Publishing Company Assoc. and John Wiley Interscience, New York, 1989, 1992). It is appreciated to one skilled in the art that the polynucleotides of SEQ ID NOS: 1-14 and 57-60 or fragments thereof (comprising at least 12 contiguous nucleotides), are particularly useful probes. Other particular useful probes for this purpose are hybridizable fragments to the sequences of SEQ ID NOS: 1-14 and 57-60 (i.e., comprising at least 12 contiguous nucleotides).

With respect to nucleic acid sequences which hybridize to specific nucleic acid sequences disclosed herein, hybridization may be carried out under conditions of reduced stringency, medium stringency or even stringent conditions. As an example of oligonucleotide hybridization, a polymer membrane containing immobilized denatured nucleic acids is first prehybridized for 30 minutes at 45 °C in a solution consisting of 0.9 M NaCl, 50 mM NaH₂PO₄, pH 7.0, 5.0 mM Na₂EDTA, 0.5% SDS, 10X Denhardt's, and 0.5 mg/ml polyriboadenylic acid. Approximately 2 X 10⁷ cpm (specific activity 4-9 X 10 cpm/ug) of ³²P end-labeled oligonucleotide probe are then added to the solution. After 12-16 hours of incubation, the membrane is washed for 30 minutes at room temperature in 1X SET (150 mM NaCl, 20 mM Tris hydrochloride, pH 7.8, 1 mM Na₂EDTA) containing 0.5% SDS, followed by a 30 minute wash in fresh 1X SET at Tm 10°C for the oligonucleotide probe. The membrane is then exposed to auto-radiographic film for detection of hybridization signals.

Stringent conditions means hybridization will occur only if there is at least 90% identity, preferably at least 95% identity and most preferably at least 97% identity between the sequences. Further, it is understood that a section of a 100 bps sequence that is 95 bps in length has 95% identity with the 1090 bps sequence from which it is obtained. See J. Sambrook et al., Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory (1989) which is hereby incorporated by reference in its entirety. Also, it is understood that a fragment of a 100 bps sequence that is 95 bps in length has 95% identity with the 100 bps sequence from which it is obtained.

As used herein, a first DNA (RNA) sequence is at least 70% and preferably at least 80% identical to another DNA (RNA) sequence if there is at least 70% and preferably at

5

10

15

20

least a 80% or 90% identity, respectively, between the bases of the first sequence and the bases of the another sequence, when properly aligned with each other, for example when aligned by BLASTN.

"Identity" as the term is used herein, refers to a polynucleotide sequence which comprises a percentage of the same bases as a reference polynucleotide (SEQ ID NOS:1-14 and 57-60). For example, a polynucleotide which is at least 90% identical to a reference polynucleotide, has polynucleotide bases which are identical in 90% of the bases which make up the reference polynucleotide and may have different bases in 10% of the bases which comprise that polynucleotide sequence.

The present invention relates polynucleotides which differ from the reference polynucleotide such that the changes are silent changes, for example the change do not alter the amino acid sequence encoded by the polynucleotide. The present invention also relates to nucleotide changes which result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference polynucleotide. In a preferred aspect of the invention these polypeptides retain the same biological action as the polypeptide encoded by the reference polynucleotide.

It is also appreciated that such probes can be and are preferably labeled with an analytically detectable reagent to facilitate identification of the probe. Useful reagents include but are not limited to radioactivity, fluorescent dyes or enzymes capable of catalyzing the formation of a detectable product. The probes are thus useful to isolate complementary copies of DNA from other sources or to screen such sources for related sequences.

The polynucleotides of this invention were recovered from genomic gene libraries from the organisms listed in Table 1. For example, gene libraries can be generated in the Lambda ZAP II cloning vector (Stratagene Cloning Systems). Mass excisions can be performed on these libraries to generate libraries in the pBluescript phagemid. Libraries are thus generated and excisions performed according to the protocols/methods hereinafter described.

5

10

15

20

The excision libraries are introduced into the *E. coli* strain BW14893 F'kan1A. Expression clones are then identified using a high temperature filter assay. Expression clones encoding several glucanases and several other glycosidases are identified and repurified. The polynucleotides, and enzymes encoded thereby, of the present invention, yield the activities as described above.

The coding sequences for the enzymes of the present invention were identified by screening the genomic DNAs prepared for the clones having glucosidase or galactosidase activity.

An example of such an assay is a high temperature filter assay wherein expression clones were identified by use of high temperature filter assays using buffer Z (see recipe below) containing 1 mg/ml of the substrate 5-bromo-4-chloro-3-indolyl-β-D-glucopyranoside (XGLU) (Diagnostic Chemicals Limited or Sigma) after introducing an excision library into the *E. coli* strain BW14893 F'kan1A. Expression clones encoding XGLUases were identified and repurified from M11TL, OC1/4V, Pyrococcus furiosus VC1, Staphylothemus marinus F1, Thermococcus 9N-2, Thermotoga maritima MSB8, Thermococcus alcaliphilus AEDII12RA, and Thermococcus chitonophagus GC74.

Z-buffer: (referenced in Miller, J.H. (1992) A Short Course in Bacterial Genetics, p. 445.)

per liter:

10

15

20

25

 $Na_2HPO_4-7H_2O$ 16.1g $NaH_2PO_4-7H_2O$ 5.5g KCl 0.75g $MgSO_4-7H_2O$ 0.246g β-mercaptoethanol 2.7ml

Adjust pH to 7.0

High Temperature Filter Assay

(1) The f factor fkan (from E. coli strain CSH118)(1) was introduced into the pho-pnh-lac-strain BW14893(2). BW13893(2). The filamentous phage library was plated on the resulting strain, BW14893 Fkan. (Miller, J.H. (1992) A Short Course in

Bacterial Genetics, Lee, K.S., Metcalf, et al., (1992) Evidence for two phosphonate degradative pathways in Enterobacter Aerogenes, J. Bacteriol., 174:2501-2510.

- (2) After growth on 100 mm LB plates containing 100 μg/ml ampicillin, 80 μg/ml nethicillin and 1mM IPTG, colony lifts were performed using Millipore HATF membrane filters.
- (3) The colonies transferred to the filters were lysed with chloroform vapor in 150 mm glass petri dishes.
- (4) The filters were transferred to 100 mm glass petri dishes containing a piece of Whatman 3MM filter paper saturated with buffer.
 - (a) when testing for galactosidase activity (XGALase), 3MM paper was saturated with Z buffer containing 1 mg/ml XGAL (ChemBridge Corporation). After transferring filter bearing lysed colonies to the glass petri dish, placed dish in oven at 80-85°C.
 - (b) when testing for glucosidase (XGLUase), 3MM paper was saturated with Z buffer containing 1 mg/ml XGLU. After transferring filter bearing lysed colonies to the glass petri dish, placed dish in oven at 80-85°C.
- (5) Positives' were observed as blue spots on the filter membranes. Used the following filter rescue technique to retrieve plasmid from lysed positive colony. Used pasteur pipette (or glass capillary tube) to core blue spots on the filter membrane. Placed the small filter disk in an Eppendorf tube containing 20 μl water. Incubated the Eppendorf tube at 75°C for 5 minutes followed by vortexing to elute plasmid DNA off filter. This DNA was transformed into electrocompetent *E. coli* cells DH10B for Thermatoga maritima MSB8-6G, Staphylothermus marinus F1-12G, Thermococcus AEDII12RA-18B/G, Thermococcus chitonophagus GC74-22G, M11Tl and OC1/4V. Electrocompetent BW14893 F'kan1A *E. coli* were used for Thermococcus 9N2-31B/G, and *Pyrococcus furiosus* VC1-7G1. Repeated filter-lift assay on transformation plates to identify 'positives'. Return transformation plates to 37°C incubator after filter lift to regenerate colonies. Inoculate 3 ml LB liquid containing 100 μg/ml ampicillin with repurified positives and incubate at 37°C

5

10

15

20

overnight. Isolate plasmid DNA from these cultures and sequence plasmid insert. In some instances where the plates used for the initial colony lifts contained non-confluent colonies, a specific colony corresponding to a blue spot on the filter could be identified on a regenerated plate and repurified directly, instead of using the filter rescue technique.

Another example of such an assay is a variation of the high temperature filter assay wherein colony-laden filters are heat-killed at different temperatures (for example, 105°C for 20 minutes) to monitor thermostability. The 3MM paper is saturated with different buffers (i.e., 100 mM NaCl, 5 mM MgCl₂, 100 mM Tris-Cl (pH 9.5)) to determine enzyme activity under different buffer conditions.

10

15

20

25

A β -glucosidase assay may also be employed, wherein Glcp β Np is used as an artificial substrate (aryl- β -glucosidase). The increase in absorbance at 405 nm as a result of p-nitrophenol (pNp) liberation was followed on a Hitachi U-1100 spectrophotometer, equipped with a thermostatted cuvette holder. The assays may be performed at 80°C or 90°C in closed 1-ml quartz cuvette. A standard reaction mixture contains 150 mM trisodium substrate, pH 5.0 (at 80°C), and 0.95 mM pNp derivative pNp = 0.561 mM⁻¹ cm⁻¹). The reaction mixture is allowed to reach the desired temperature, after which the reaction is started by injecting an appropriate amount of enzyme (1.06 ml final volume).

1 U β -glucosidase activity is defined as that amount required to catalyze the formation of 1.0 μ mol pNp/min. D-cellobiose may also be used as a substrate.

An ONPG assay for β -galactosidase activity is described by Miller, J.H. (1992) A Short Course in Bacterial Genetics and Mill, J.H. (1992) Experiments in Molecular Genetics, the contents of which are hereby incorporated by reference in their entirety.

A quantitative fluorometric assay for β -galactosidase specific activity is described by : Youngman P., (1987) Plasmid Vectors for Recovering and Exploiting Tn917 Transpositions in Bacillus and other Gram-Positive Bacteria. In Plasmids: A Practical approach (ed. K. Hardy) pp 79-103. IRL Press, Oxford. A description of the procedure can be found in Miller (1992) p. 75-77, the contents of which are incorporated by reference herein in their entirety.

The polynucleotides of the present invention may be in the form of DNA which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequences which encodes the mature enzymes may be identical to the coding sequences shown in Figures 1-8 (SEQ ID NOS: 1-14 and 57-60) or may be a different coding sequence which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same mature enzymes as the DNA of Figures 1-18 (SEQ ID NOS: 1-14 and 57-60).

The polynucleotide which encodes for the mature enzyme of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) may include, but is not limited to: only the coding sequence for the mature enzyme; the coding sequence for the mature enzyme and additional coding sequence such as a leader sequence or a proprotein sequence; the coding sequence for the mature enzyme (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequence 5' and/or 3' of the coding sequence for the mature enzyme.

Thus, the term "polynucleotide encoding an enzyme (protein)" encompasses a polynucleotide which includes only coding sequence for the enzyme as well as a polynucleotide which includes additional coding and/or non-coding sequence.

The present invention further relates to variants of the hereinabove described polynucleotides which encode for fragments, analogs and derivatives of the enzymes having the deduced amino acid sequences of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64). The variant of the polynucleotide may be a naturally occurring allelic variant of the polynucleotide or a non-naturally occurring variant of the polynucleotide.

Thus, the present invention includes polynucleotides encoding the same mature enzymes as shown in Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) as well as variants of such polynucleotides which variants encode for a fragment, derivative or analog of the enzymes of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64). Such nucleotide variants include deletion variants, substitution variants and addition or insertion variants.

As hereinabove indicated, the polynucleotides may have a coding sequence which is a naturally occurring allelic variant of the coding sequences shown in Figures 1-18 (SEQ

5

10

15

20

ID NOS: 1-14 and 57-60). As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded enzyme.

. 5

15

20

25

Fragments of the full length gene of the present invention may be used as a hybridization probe for a cDNA or a genomic library to isolate the full length DNA and to isolate other DNAs which have a high sequence similarity to the gene or similar biological activity. Probes of this type preferably have at least 10, preferably at least 15, and even more preferably at least 30 bases and may contain, for example, at least 50 or more bases. The probe may also be used to identify a DNA clone corresponding to a full length transcript and a genomic clone or clones that contain the complete gene including regulatory and promotor regions, exons, and introns. An example of a screen comprises isolating the coding region of the gene by using the known DNA sequence to synthesize an oligonucleotide probe. Labeled oligonucleotides having a sequence complementary to that of the gene of the present invention are used to screen a library of genomic DNA to determine which members of the library the probe hybridizes to.

The present invention further relates to polynucleotides which hybridize to the hereinabove-described sequences if there is at least 70%, preferably at least 90%, and more preferably at least 95% identity between the sequences. The present invention particularly relates to polynucleotides which hybridize under stringent conditions to the hereinabove-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. The polynucleotides which hybridize to the hereinabove described polynucleotides in a preferred embodiment encode enzymes which either retain substantially the same biological function or activity as the mature enzyme encoded by the DNA of Figures 1-18 (SEQ ID NOS: 1-14 and 57-60).

Alternatively, the polynucleotide may have at least 15 bases, preferably at least 30 bases, and more preferably at least 50 bases which hybridize to any part of a polynucleotide of the present invention and which has an identity thereto, as hereinabove described, and which may or may not retain activity. For example, such polynucleotides may be employed

as probes for the polynucleotides of SEQ ID NOS: 1-14 and 57-60, for example, for recovery of the polynucleotide or as a diagnostic probe or as a PCR primer.

Thus, the present invention is directed to polynucleotides having at least a 70% identity, preferably at least 90% identity and more preferably at least a 95% identity to a polynucleotide which encodes the enzymes of SEQ ID NOS: 15-28 and 61-64 as well as fragments thereof, which fragments have at least 15 bases, preferably at least 30 bases and most preferably at least 50 bases, which fragments are at least 90% identical, preferably at least 95% identical and most preferably at least 97% identical under stringent conditions to any portion of a polynucleotide of the present invention.

The present invention further relates to enzymes which have the deduced amino acid sequences of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) as well as fragments, analogs and derivatives of such enzyme.

The terms "fragment," "derivative" and "analog" when referring to the enzymes of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) means enzymes which retain essentially the same biological function or activity as such enzymes. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature enzyme.

The enzymes of the present invention may be a recombinant enzyme, a natural enzyme or a synthetic enzyme, preferably a recombinant enzyme.

The fragment, derivative or analog of the enzymes of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature enzyme is fused with another compound, such as a compound to increase the half-life of the enzyme (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature enzyme, such as a leader or secretory sequence or a sequence which is employed for purification of the mature enzyme or a proprotein sequence. Such fragments, derivatives

5

10

15

20

and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

The enzymes and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or enzyme present in a living animal is not isolated, but the same polynucleotide or enzyme, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or enzymes could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

The enzymes of the present invention include the enzymes of SEQ ID NOS: 15-28 and 61-64 (in particular the mature enzyme) as well as enzymes which have at least 70% similarity (preferably at least 70% identity) to the enzymes of SEQ ID NOS: 15-28 and 61-64 and more preferably at least 90% similarity (more preferably at least 90% identity) to the enzymes of SEQ ID NOS: 15-28 and 61-64 and still more preferably at least 95% similarity (still more preferably at least 95% identity) to the enzymes of SEQ ID NOS: 15-28 and 61-64 and also include portions of such enzymes with such portion of the enzyme generally containing at least 30 amino acids and more preferably at least 50 amino acids.

As known in the art "similarity" between two enzymes is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one enzyme to the sequence of a second enzyme.

A variant, i.e. a "fragment", "analog" or "derivative" polypeptide, and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.

Among preferred variants are those that vary from a reference by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala,

5

10

15

20

Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

Most highly preferred are variants which retain the same biological function and activity as the reference polypeptide from which it varies.

Fragments or portions of the enzymes of the present invention may be employed for producing the corresponding full-length enzyme by peptide synthesis: therefore, the fragments may be employed as intermediates for producing the full-length enzymes. Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of enzymes of the invention by recombinant techniques.

Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the present invention. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

The polynucleotides of the present invention may be employed for producing enzymes by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety of expression vectors for expressing an enzyme. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

5

10

15

20

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the <u>E. colilac</u> or <u>trp</u>, the phage lambda P_L promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

10

15

20

25

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in <u>E. coli</u>.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as <u>E. coli</u>, <u>Streptomvces</u>, <u>Bacillus subtilis</u>; fungal cells, such as yeast: insect cells such as <u>Drosophila S2</u> and <u>Spodoptera Sf9</u>; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and

promoters are known to those of skill in the art. and are commercially available. The following vectors are provided by way of example; Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pD10, psiX174, pBluescript II KS, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); Eukaryotic: pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, P_L and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the enzymes of the invention can be synthetically produced by conventional peptide synthesizers.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., Molecular Cloning: A Laboratory

5

10

15

20

Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Transcription of the DNA encoding the enzymes of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Examples include the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

5

10

15

20

25

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of E. coli and S. cerevisiae TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated enzyme. Optionally, the heterologous sequence can encode a fusion enzyme including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include <u>E. coli</u>, <u>Bacillus subtilis</u>, <u>Salmonella typhimurium</u> and various species within the genera Pseudomonas, Streptomyces, and Staphylococcus, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from

commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well known to those skilled in the art.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell, 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

The enzyme can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing

10

15

20

configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The enzymes of the present invention may be a naturally purified product, or a product of chemical synthetic procedures, or produced by recombinant techniques from a prokaryotic or eukaryotic host (for example, by bacterial, yeast, higher plant, insect and mammalian cells in culture). Depending upon the host employed in a recombinant production procedure, the enzymes of the present invention may be glycosylated or may be non-glycosylated. Enzymes of the invention may or may not also include an initial methionine amino acid residue.

 β -galactosidase hydrolyzes lactose to galactose and glucose. Accordingly, the OC1/4V, 9N2-31B/G, AEDII12RA-18B/G and F1-12G enzymes may be employed in the food processing industry for the production of low lactose content milk and for the production of galactose or glucose from lactose contained in whey obtained in a large amount as a by-product in the production of cheese. Generally, it is desired that enzymes used in food processing, such as the aforementioned β -galactosidases, be stable at elevated temperatures to help prevent microbial contamination.

10

15

20

25

These enzymes may also be employed in the pharmaceutical industry. The enzymes are used to treat intolerance to lactose. In this case, a thermostable enzyme is desired, as well. Thermostable β -galactosidases also have uses in diagnostic applications, where they are employed as reporter molecules.

Glucosidases act on soluble cellooligosaccharides from the non-reducing end to give glucose as the sole product. Glucanases (endo- and exo-) act in the depolymerization of cellulose, generating more non-reducing ends (endo-glucanases, for instance, act on internal linkages yielding cellobiose, glucose and cellooligosaccharides as products). β-glucosidases are used in applications where glucose is the desired product. Accordingly, M11TL, F1-12G, GC74-22G, MSB8-6G, OC1/4V, VC1-7G1, 9N2-31B/G and AEDII12RA18B/G may be employed in a wide variety of industrial applications, including in corn wet milling for the separation of starch and gluten, in the fruit industry for clarification and equipment maintenance, in baking for viscosity reduction, in the textile

industry for the processing of blue jeans, and in the detergent industry as an additive. For these and other applications, thermostable enzymes are desirable.

Antibodies generated against the enzymes corresponding to a sequence of the present invention can be obtained by direct injection of the enzymes into an animal or by administering the enzymes to an animal, preferably a nonhuman. The antibody so obtained will then bind the enzymes itself. In this manner, even a sequence encoding only a fragment of the enzymes can be used to generate antibodies binding the whole native enzymes. Such antibodies can then be used to isolate the enzyme from cells expressing that enzyme.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, 1975, Nature, 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic enzyme products of this invention. Also, transgenic mice may be used to express humanized antibodies to immunogenic enzyme products of this invention.

Antibodies generated against the enzyme of the present invention may be used in screening for similar enzymes from other organisms and samples. Such screening techniques are known in the art, for example, one such screening assay is described in "Methods for Measuring Cellulase Activities", *Methods in enzymology*, Vol 160, pp. 87-116, which is hereby incorporated by reference in its entirety.

The present invention will be further described with reference to the following examples; however, it is to be understood that the present invention is not limited to such examples. All parts or amounts, unless otherwise specified, are by weight.

In order to facilitate understanding of the following examples certain frequently occurring methods and/or terms will be described.

10

15

20

"Plasmids" are designated by a lower case p preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

"Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1 µg of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 µl of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 µg of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a polyacrylamide gel to isolate the desired fragment.

10

15

20

25

Size separation of the cleaved fragments is performed using 8 percent polyacrylamide gel described by Goeddel, D. et al., Nucleic Acids Res., 8:4057 (1980).

"Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T., et al., Id., p. 146). Unless otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units of T4 DNA ligase ("ligase") per 0.5 µg of approximately equimolar amounts of the DNA fragments to be ligated.

Unless otherwise stated, transformation was performed as described in the method of Graham, F. and Van der Eb, A., Virology, 52:456-457 (1973).

Example 1

Bacterial Expression and Purification of Glycosidase Enzymes

DNA encoding the enzymes of the present invention, SEQ ID NOS: 1-14 and 57-60 were initially amplified from a pBluescript vector containing the DNA by the PCR technique using the primers noted herein. The amplified sequences were then inserted into the respective PQE vector listed beneath the primer sequences, and the enzyme was expressed according to the protocols set forth herein. The 5' and 3' primer sequences for the respective genes are as follows:

Thermococcus AEDII12RA -18B/G

5' CCGAGAATTCATTAAAGAGGAGAAATTAACTATGGTGAATGCTATGATTGTC 3' (SEQ ID NO:29)

3' CGGAAGATCTTCATAGCTCCGGAAGCCCATA 5' (SEQ ID NO:30)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Blg II.

OC1/4V-33B/G

5' CCGAGAATTCATTAAAGAGGGGAGAAATTAACTATGATAAGAAGGTCCGATTTTCC 3' (SEQ ID NO:31)

3' CGGAAGATCTTTAAGATTTTAGAAATTCCTT 5' (SEQ ID NO:32)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Bgl II.

Thermococcus 9N2 - 31B/G

5' CCGAGAATTCATTAAAGAGGAGAAATTAACTATGCTACCAGAAGGCTTTCTC 3' (SEQ ID NO:33)

3' CGGAGGTACCTCACCCAAGTCCGAACTTCTC 5' (SEQ ID NO:34)

Vector: pQE30; and contains the following restriction enzyme sites 5' EcoRI and 3' Kpnl.

10

15

20

Staphylothermus marinus F1 - 12G

5' CCGAGAATTCATTAAAGAGGAGAAATTAACTATGATAAGGTTTCCTGATTAT 3' (SEO ID NO:35)

3' CGGAAGATCTTTATTCGAGGTTCTTTAATCC 5' (SEQ ID NO:36)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Bgl II.

Thermococcus chitonophagus GC74 - 22G

5' CCGAGAATTCATTAAAGAGGGAGAAATTAACTATGCTTCCAGGAGAACTTTCTC 3' (SEQ ID NO:37)

3' CGGAGGATCCCTACCCCTCTAAGATCTC 5' (SEQ ID NO:38)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' BamHI.

M11TL

10

15

20

25

5' AATAATCTAGAGCATGCAATTCCCCAAAGACTTCATGATAG 3' (SEQ ID NO:39)

3' AATAAAAGCTTACTGGATCAGTGTAAGATGCT 5' (SEQ ID NO:40)

Vector: pQE70; and contains the following restriction enzyme sites 5' SphI and 3' Hind III.

Thermotoga maritima MSB8-6G

5' CCGACAATTGATTAAAGAGGAGAAATTAACTATGGAAAGGATCGATGAAATT 3' (SEQ ID NO:41)

3' CGGAGGTACCTCATGGTTTGAATCTCTTCTC 5' (SEQ ID NO:42)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' KpnI.

Pyrococcus furiosus VC1 - 7G1

5' CCGAGAATTGATTAAAGAGGAGAAATTAACTATGTTCCCTGAAAAGTTCCTT 3' (SEQ ID NO:43)

3' CGGAGGTACCTCATCCCCTCAGCAATTCCTC 5' (SEQ ID NO:44)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Kpn I.

Bankia gouldi endoglucanase (37GP1)

- 5' AATAAGGATCCGTTTAGCGACGCTCGC 3' (SEQ ID NO:45)
- 3' AATAAAAGCTTCCGGGTTGTACAGCGGTAATAGGC 5' (SEQ ID NO:46)

Vector: pQE52; and contains the following restriction enzyme sites 5' Bam HI and 3' Hind III.

Thermotoga maritima α-galactosidase (6GC2)

- 5' TTTATTGAATTCATTAAAGAGGAGAAATTAACTATGATCTGTGTGGAAATATTCGGAAAG 3' (SEQ ID NO:47)
- 3' TCTATAAAGCTTTCATTCTCTCACCCTCTTCGTAGAAG 5' (SEQ ID NO:48)

Vector: pQET; and contains the following restriction enzyme sites 5' EcoRI and 3' Hind III.

Thermotoga maritima \(\beta\)-mannanase (6GP2)

- 5' TTTATTCAATTGATTAAAGAGGAGAAATTAACTATGGGGATTGGTGGCGACGAC 3' (SEQ ID NO:49)
- 3' TTTATTAAGCTTATCTTTTCATATTCACATACCTCC 5' (SEQ ID NO:50)

Vector: pQEt; and contains the following restriction enzyme sites 5' Hind III and 3' EcoRI.

AEPII 1a \(\beta\)-mannanase (63GB1)

- 5' TTTATTGAATTCATTAAAGAGGAGAAATTAACTATGCTACCAGAAGAGTTCCTATGGGGC 3' (SEQ ID NO:51)
- 3' TTTATTAAGCTTCTCATCAACGGCTATGGTCTTCATTTC 5' (SEQ ID NO:52)

Vector: pQEt; and contains the following restriction enzyme sites 5' Hind III and 3' EcoRI.

OC1/4V endoglucanase (33GP1)

- 5' AAAAAACAATTGAATTCATTAAAGAGGAGAAATTAACTATGGTAGAAAGACACTTCAGATATGTTCTT
 3' (SEQ ID NO:53)
 - 3' TTTTTCGGATCCAATTCTTCATTTACTCTTTGCCTG 5' (SEQ ID NO:54)

5 .

łO

15

20

Vector: pQEt; and contains the following restriction enzyme sites 5' BamHI and 3' EcoRI.

Thermotoga maritima pullalanase (6GP3)
5' TTTTGGAATTCATTAAAGAGGAGAAATTAACTATGGAACTGATCATAGAAGGTTAC 3'
(SEQ ID NO:55)
3' ATAAGAAGCTTTTCACTCTCTGTACAGAACGTACGC 5' (SEQ ID NO:56)
Vector: pQEt; and contains the following restriction enzyme sites 5' EcoRI and 3' Hind III.

The restriction enzyme sites indicated correspond to the restriction enzyme sites on the bacterial expression vector indicated for the respective gene (Qiagen, Inc. Chatsworth, CA). The pQE vector encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulatable promoter operator (P/O), a ribosome binding site (RBS), a 6-His tag and restriction enzyme sites.

10

15

20

25

The pQE vector was digested with the restriction enzymes indicated. The amplified sequences were ligated into the respective pQE vector and inserted in frame with the sequence encoding for the RBS. The ligation mixture was then used to transform the E. coli strain M15/pREP4 (Qiagen, Inc.) by electroporation. M15/pREP4 contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants were identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies were selected. Plasmid DNA was isolated and confirmed by restriction analysis. Clones containing the desired constructs were grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture was used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells were grown to an optical density 600 (O.D.600) of between 0.4 and IPTG ("Isopropyl-B-D-thiogalacto pyranoside") was then added to a final 0.6. concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression. Cells were grown an extra 3 to 4 hours. Cells were then harvested by centrifugation.

The primer sequences set out above may also be employed to isolate the target gene from the deposited material by hybridization techniques described above.

Example 2

Isolation of A Selected Clone From the Deposited genomic clones

A clone is isolated directly by screening the deposited material using the oligonucleotide primers set forth in Example 1 for the particular gene desired to be isolated. The specific oligonucleotides are synthesized using an Applied Biosystems DNA synthesizer. The oligonucleotides are labeled with 32P--ATP using T4 polynucleotide kinase and purified according to a standard protocol (Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY, 1982). The deposited clones in the pBluescript vectors may be employed to transform bacterial hosts which are then plated on 1.5% agar plates to the density of 20,000-50,000 pfu/150 mm plate. These plates are screened using Nylon membranes according to the standard screening protocol (Stratagene, 1993). Specifically, the Nylon membrane with denatured and fixed DNA is prehybridized in 6 x SSC, 20 mM NaH₂PO₄, 0.4%SDS, 5 x Denhardt's 500 μg/ml denatured, sonicated salmon sperm DNA; and 6 x SSC, 0.1% SDS. After one hour of prehybridization, the membrane is hybridized with hybridization buffer 6xSSC, 20 mM NaH2PO4, 0.4%SDS. 500 ug/ml denatured, sonicated salmon sperm DNA with 1x106 cpm/ml 32P-probe overnight at 42°C. The membrane is washed at 45-50°C with washing buffer 6 x SSC, 0.1% SDS for 20-30 minutes dried and exposed to Kodak X-ray film overnight. Positive clones are isolated and purified by secondary and tertiary screening. The purified clone is sequenced to verify its identity to the primer sequence.

Once the clone is isolated, the two oligonucleotide primers corresponding to the gene of interest are used to amplify the gene from the deposited material. A polymerase chain reaction is carried out in 25 μ l of reaction mixture with 0.5 ug of the DNA of the gene of interest. The reaction mixture is 1.5-5 mM MgCl₂, 0.01% (w/v) gelatin, 20 μ M each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq

5

15

20

WO 98/24799 PCT/US97/22623

polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with the Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the gene of interest by subcloning and sequencing the DNA product. The ends of the newly purified genes are nucleotide sequenced to identify full length sequences. Complete sequencing of full length genes is then performed by Exonuclease III digestion or primer walking.

Example 3

Screening for Galactosidase Activity

Screening procedures for α -galactosidase protein activity may be assayed for as follows:

10

15

20

Substrate plates were provided by a standard plating procedure. Dilute XL1-Blue MRF $E\ coli$ host of (Stratagene Cloning Systems, La Jolla, CA) to O.D. $_{600}$ = 1.0 with NZY media. In 15 ml tubes, inoculate 200 μ l diluted host cells with phage. Mix gently and incubate tubes at 37 °C for 15 min. Add approximately 3.5 ml LB top agarose (0.7%) containing 1mM IPTG to each tube and pour onto all NYZ plate surface. Allow to cool and incubate at 37 °C overnight. The assay plates are obtained as substrate p-Nitrophenyl α -galactosidase (Sigma) (200 mg/100 ml) (100 mM NaCl, 100 mM Potassium-Phosphate) 1% (w/v) agarose. The plaques are overlayed with nitrocellulose and incubated at 4 °C for 30 minutes whereupon the nitrocellulose is removed and overlayed onto the substrate plates. The substrate plates are then incubated at 70 °C for 20 minutes.

WO 98/24799 PCT/US97/22623

Example 4

Screening of Clones for Mannanase Activity

A solid phase screening assay was utilized as a primary screening method to test clones for \(\beta \)-mannanase activity.

A culture solution of the Y1090-E. coli host strain (Stratagene Cloning Systems, La Jolla, CA) was diluted to O.D.₆₀₀=1.0 with NZY media. The amplified library from Thermotoga maritima lambda gtl1 library was diluted in SM (phage dilution buffer): 5×10^7 pfu/µl diluted 1:1000 then 1:100 to 5×10^2 pfu/µl. Then 8 µl of phage dilution (5×10^2 pfu/µl) was plated in 200 µl host cells. They were then incubated in 15 ml tubes at 37 °C for 15 minutes.

Approximately 4 ml of molten, LB top agarose (0.7%) at approximately 52 °C was added to each tube and the mixture was poured onto the surface of LB agar plates. The agar plates were then incubated at 37 °C for five hours. The plates were replicated and induced with 10 mM IPTG-soaked Duralon-UVTM nylon membranes (Stratagene Cloning Systems, La Jolla, CA) overnight. The nylon membranes and plates were marked with a needle to keep their orientation and the nylon membranes were then removed and stored at 4 °C.

An Azo-galactomannan overlay was applied to the LB plates containing the lambda plaques. The overlay contains 1% agarose, 50 mM potassium-phosphate buffer pH 7, 0.4% Azocarob-galactomannan. (Megazyme, Australia). The plates were incubated at 72 °C. The Azocarob-galactomannan treated plates were observed after 4 hours then returned to incubation overnight. Putative positives were identified by clearing zones on the Azocarob-galactomannan plates. Two positive clones were observed.

The nylon membranes referred to above, which correspond to the positive clones were retrieved, oriented over the plate and the portions matching the locations of the clearing zones for positive clones were cut out. Phage was eluted from the membrane cut-out portions by soaking the individual portions in 500 µl SM (phage dilution buffer) and 25 µl CHCl₃.

10

15

20

25

Example 5

Screening of Clones for Mannosidase Activity

A solid phase screening assay was utilized as a primary screening method to test clones for \(\beta \)-mannosidase activity.

5

10

15

20

25

A culture solution of the Y1090-*E. coli* host strain (Stratagene Cloning Systems, La Jolla, CA) was diluted to O.D. $_{600}$ =1.0 with NZY media. The amplified library from AEPII 1a lambda gtl1 library was diluted in SM (phage dilution buffer): 5×10^7 pfu/µl diluted 1:1000 then 1:100 to 5×10^2 pfu/µl. Then 8 µl of phage dilution (5×10^2 pfu/µl) was plated in 200 µl host cells. They were then incubated in 15 ml tubes at 37 °C for 15 minutes.

Approximately 4 ml of molten, LB top agarose (0.7%) at approximately 52 °C was added to each tube and the mixture was poured onto the surface of LB agar plates. The agar plates were then incubated at 37 °C for five hours. The plates were replicated and induced with 10 mM IPTG-soaked Duralon-UVTM nylon membranes (Stratagene Cloning Systems, La Jolla, CA) overnight. The nylon membranes and plates were marked with a needle to keep their orientation and the nylon membranes were then removed and stored at 4 °C.

A p-nitrophenyl-\(\beta\)-D-manno-pyranoside overlay was applied to the LB plates containing the lambda plaques. The overlay contains 1% agarose, 50 mM potassium-phosphate buffer pH 7, 0.4% p-nitrophenyl-\(\beta\)-D-manno-pyranoside. (Megazyme, Australia). The plates were incubated at 72 °C. The p-nitrophenyl-\(\beta\)-D-manno-pyranoside treated plates were observed after 4 hours then returned to incubation overnight. Putative positives were identified by clearing zones on the p-nitrophenyl-\(\beta\)-D-manno-pyranoside plates. Two positive clones were observed.

The nylon membranes referred to above, which correspond to the positive clones were retrieved, oriented over the plate and the portions matching the locations of the clearing zones for positive clones were cut out. Phage was eluted from the membrane cut-out portions by soaking the individual portions in 500 µl SM (phage dilution buffer) and 25 µl CHCl₃.

5

10

15

20

25

Example 6

Screening for Pullulanase Activity

Screening procedures for pullulanase protein activity may be assayed for as follows:

Substrate plates were provided by a standard plating procedure. Host cells are diluted to O.D. $_{600}$ = 1.0 with NZY or appropriate media. In 15 ml tubes, inoculate 200 μ l diluted host cells with phage. Mix gently and incubate tubes at 37 °C for 15 min. Add approximately 3.5 ml LB top agarose (0.7%) is added to each tube and the mixture is plated, allowed to cool, and incubated at 37 °C for about 28 hours. Overlays of 4.5 mls of the following substrate are poured:

100 ml total volume

0.5g	Red Pullulan Red (Megazyme, Australia)
1.0g	Agarose
5ml	Buffer (Tris-HCL pH 7.2 @ 75 °C)
2ml	5M NaCl
5ml	CaCl ₂ (100mM)
85ml	dH₂O

Plates are cooled at room temperature, and thenm incubated at 75°C for 2 hours. Positives are observed as showing substrate degradation.

Example 7

Screening for Endoglucanase Activity

Screening procedures for endoglucanase protein activity may be assayed for as follows:

1. The gene library is plated onto 6 LB/GelRite/0.1% CMC/NZY agar plates (~4,800 plaque forming units/plate) in E.coli host with LB agarose as top agarose. The plates are incubated at 37°C overnight.

- 2. Plates are chilled at 4°C for one hour.
- 3. The plates are overlayed with Duralon membranes (Stratagene) at room temperature for one hour and the membranes are oriented and lifted off the plates and stored at 4°C.
- 4. The top agarose layer is removed and plates are incubated at 37°C for ~3 hours.
 - 5. The plate surface is rinsed with NaCl.
 - 6. The plate is stained with 0.1% Congo Red for 15 minutes.
 - 7. The plate is destained with 1M NaCl.

10

15

20

25

- 8. The putative positives identified on plate are isolated from the Duralon membrane (positives are identified by clearing zones around clones). The phage is eluted from the membrane by incubating in $500\mu l$ SM + $25\mu l$ CHCl₃ to elute.
- 9. Insert DNA is subcloned into any appropriate cloning vector and subclones are reassayed for CMCase activity using the following protocol:
- i) Spin 1ml overnight miniprep of clone at maximum speed for 3 minutes.
- ii) Decant the supernatant and use it to fill "wells" that have been made in an LB/GelRite/0.1% CMC plate.
 - iii) Incubate at 37°C for 2 hours.
 - iv) Stain with 0.1% Congo Red for 15 minutes.
 - v) Destain with 1M NaCl for 15 minutes.
 - vi) Identify positives by clearing zone around clone.

Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, within the scope of the appended claims, the invention may be practiced otherwise than as particularly described.

WO 98/24799 PCT/US97/22623

WHAT IS CLAIMED IS:

1. An isolated polynucleotide selected from the group consisting of:

- (a) SEQ ID NOS: 1-14 and 57-60;
- (b) SEQ ID NOS: 1-14 and 57-60, wherein T can also be U;
- (c) polynucleotide sequences complementary to SEQ ID NOS: 1-14 and 57-60;
- (d) polynucleotide sequences which encode an amino acid sequence as set forth in SEQ ID NOS:15-28, and 61-64; and
- (e) fragments of (a), (b), (c) or (d) that are at least 15 consecutive bases in length and that will selectively hybridize to DNA which encodes a polypeptide of SEQ ID NOS:15-28, and 61-64.
- 2. A vector comprising a polynucleotide of claim 1.
- 3. A host cell containing the vector of claim 2.
- 4. The method of claim 3, wherein the host cell is a eukaryotic cell.
- 5. The method of claim 3, wherein the host cell is a prokaryotic cell.
- 6. A method for producing a polypeptide comprising:
 - (a) culturing the host cells of claim 3;
 - (b) expressing from the host cell of claim 3 a polypeptide encoded by said polynucleotide; and
 - (c) isolating the polypeptide.

WO 98/24799 PCT/US97/22623

- 7. An enzyme selected from the group consisting of:
 - (a) an enzyme comprising an amino acid sequence set forth in SEQ ID NOS: 15-28 or 61-64; and
 - (b) an enzyme which comprises at least 30 consecutive amino acid residue as an enzyme of (a).
- 8. An enzyme of which at least a portion is coded for by a polynucleotide of claim 1, and which is selected from the group consisting of:
 - (a) an enzyme comprising an amino acid sequence which is at least 70% identical to an amino acid sequence selected from the group of amino acid sequences set forth in SEQ ID NOS:15-28 or 61-64; and
 - (b) an enzyme which comprises at least 30 amino acid residues to the enzyme of (a).
- 9. A method for generating glucose from soluble cell oligosaccharides comprising contacting a sample containing oligosaccharides with an effective amount of an enyzme selected from the group consisting of an enzyme having the amino acid sequence set forth in SEQ ID NOS: 15-28, 61-63 and 64 such that glucose is produced.
- 10. The method of cliam 9, wherein the sample is selected from the group consisting of dairy products, fruit juices, detergents, textiles, guar gum, animal feed, plant biomass and waste products.
- The method of claim 9, wherein the oligosaccharide is selected from the group consisting of maltose, cellobiose, lactose, sucrose, raffinose, stachyose, verbascose, cellulose, starch, amylose, glycogen, disacharrides, polysacharrides and pullulan.

M11TL GLYCOSIDASE - 29G COMPLETE GENE SEQUENCE - 9/95

3200ENCE - 9/95	
THE AAA PTI THE AAA CAL THE ATA ATA THE TO THE TOTAL	
1 THE AAA PRO CITY AAA CAC THE ACT THE GIV TAC TEA THE THA CITY THE CAA THE CA	7' 60
21 Gly 11e Pro Gly Ser Glu Asp Pro Ash Ser Asp Trp Trp Val Trp Val His Asp Pro Gl	1.20
121 AAC ACA COA COA	10
121 AND ACA GCA GCT EGA CTA CTC ACC CER GAT TIT CCT GAG AND GCC CCA GCT TAC TER AND 41 AND THE ALE GLY Lew Val. Ser Gly AND Phe Pro Gly AND CCC CCA GCT TAC TER AND	
41 Ann Thi Ala Ala Gly Leu Val Ser Gly And Phe Pro Glu Ann Gly Pro Gly Tyr Tip And	r 180
TO AND CAR	
18) TTA AAC CAA AAT GAC CAC GAC CTG GCT GAG AAG CTG GGG GTT AAC ACT ATT AGA CTA GCG	
241 CTT GAG TGG AGT AGG ATT TTT CCA AAG CCA ACT TTC AAT GTT AAA GTC CCT GTA GAG AGA 81 Val Glu Trp Ser Arg Ile Phe Pro Lys Pro Thr Phe Asp Val Lys No. 1	300
101 ASP Glu ASR Gly Ser Ile Val His Val ASP Val ASP Lys Ala Val Glu Arg Leu ASP	360
The same of the sa	
121 Glu Leu Ala Asn Lys Glu Ala Val Asn His Tyr Val Glu Het Tyr Lys Asp Trp Val Glu	. 420
421 AGA GGT AGA ANA GOT AGA AN	140 .
421 AGA GGT AGA AMA CTT ATA CTC AMT ITA TAC CAT TGG CCC CTG CCT CTC TGG CTT CAC AMC	
The Let Tro Let Vie have	480
TOT CON ATC ATC GTG ACA ACA ACC ACC	160
161 Pro Ile Het Val Arg Arg Het Gly Pro Asp Arg Ala Pro Ser Gly Trp Leu Asn Glu Glu	. 540
To ser Gly Trp Leu Asp Clu Clu	180
THE GTG GTG GAG TTT CCC ALL THE COLUMN THE C	
181 Ser Val Val Glu Phe Ala Lys Tyr Ala Ala Tyr Ile Ala Trp Lys Het Gly Glu Leu Pro	600
Lys Met Gly Glu Leu Pro	200
	•
201 Val Het Trp Ser Thr Het Asn Glu Pro Asn Val Val Tyr Glu Gln Gly Tyr Het Phe Val	660
661 AAA GGG GGT TTC CC2 CC2 CC2 CC2 CC2 CC2 CC2 CC2 CC	220
	720
DEB GIR ALE ALE ASP LYS Ale AND AND AND	240
741 ATC ATC CAG GCT CAT GCA GCG GCG GCG	
241 Met Ile Gin Ala His Ala Arg Ala Tyr Asp Asn Ile Lys Arg Phe Ser Lys Lys Pro Val	780
The Lys Arg Phe Ser Lys Lys Pro Val	260
101 GGA CTA ATA TAC CCT TTC CAA TOO TOO	
261 Gly Leu Ile Tyr Ala Phe Gln Trp Phe Glu Leu Leu Glu Gly Pro Ala Glu Val Phe Asp	840
841 ANG TOTAL AND	280
281 Lys Phe Lys Ser Ser Lys Leu Tyr Tyr Phe Thr Asp 11e Val Car Add GGT AGT TCA ATC	
THE VALUE OF CAP TIE	900
901 ATC AAT GTT GAA TAC AGG AGA GAT CTT GCC AAT AGG CTA GAC TGG TTG GGC GTT AAC TAC 301 Ile Asn Val Glu Tyr Arg Arg Asp Leu Ala Asn Arg Leu Aco Tac	300
301 The Ash Val Glu TVY Are her car car good and age cra gac tog tre ggc gtt aac tac	960
The second of th	320
701 TAT AGE CGT TTA CTC TO SALE AND	
321 Tyr Ser Arg Leu Val Tyr Lys Ile Val Arm Arc Lys Arg Lys Ile Val Arm Arc CTG CAC GGG TAT GGA	1020
The Lie Leu His Cly The Cly	340
1021 TTC CTT TCT ACA CCT COT COT	
	1080
1081 GAG GTG TAT CCT CAN AND AND AND AND AND AND AND AND AND A	360
	1140
	1140 380
1141 GAC TIG ATC GTG ACC CAC AAG COTT	
1141 GAC TTG ATC GTG ACC GAG AAC GGT CTT TCA GAC AGC AGC GGT GGG TTG AGA CCG GCA TAC 381 ASP Leu 11c Val Thr Glu Ash Gly Val Ser Ash S	1200
The ser will have a ser and the ser are pro all a Tvi	100
1201 CTG GTC TCG CAT CEPT TIAC AND GT	
401 Lety Val Ser His Val Tyl Ser Val Typ Lys Ala Ala Ann Glu Gly tie Pro Val Lys Gly	260
and the the tree value of the tree value of the tree tree value of th	120
1201 TAC CTC CAC TEE: ACC MANY ACC	
	120
The Arg Clu Lys Ph. 4	40

Figure la

	, 11771 ·	(7);	CAC:	A74.	,	 		•	 4	'7'	1 .6 16 1	۸٠٠,	r,		۸۱.,	1.041		1 440 460 1440
1441 481	1 VC:	TAA	14	46			•		,,	.,,,,	1,4-13	Gin.	ff 1 ::	1.291	T711	laren	110	4110

Figure 1b(Continued)

OC1/4 GLYCOSIDASE - J3G/B COMPLETE GENE SEQUENCE - 9/95

COMPLETE GENE SEQUENCE - 9/95
ATT: ATA AGA AGG TCC GAT TTT CTA AAA CAT TTT ATC TTC GGA ACT: CRT ACT: CRT ACC GCA TAC 60 61 CAG ATT GAA GGT GCA ACC CAA CAG CAG CAG CAG TAC 60 61 CAG ATT GAA GGT GCA ACC CAA CAG CAG CAG CAG CAG CAG CAG
At the state of th
61 CAG ATT GAA GGT GCA GCA AAC GAA GAT GGC AGA GGG CCA TCA ATT TGG GAT GTC TTT TCA 120
Gin Tie Glu Gly Ala Ala Asa (Illy Asa) Cly AGA GGG CCA TCA ATT TGG GAT CTC
21 Gin Tie Giu Gly Ale Ale Ash Giu Asp Giy Arg Gly Pro Ser Tie Trp Asp Val Phe Ser 40
41 His The Ser 40
121 CAC ACG CCT GGC AMA ACC CTG AMG GGT GAC ACA GGA GAC GTT GCG TGT GAC CAT TAT CAC 180 181 CGA TAC AAG GAA GAT ATC CAG GTC GTC GAC GTA GAS ASP HIS TYF HIS 60
18) CGA TAC AND CIT OF THE CITY ASP THE CITY ASP WAI Ala CYS ASP HIS TYP HIS
61 AND THE LIVE CON THE CAG CTC ATC AND GAS ATA COO TO
181 CGA TAC AAG GAA GAT ATC CAG CTG ATG AAA GAA ATA GGG TTA GAC CCT TAC AGG TTC TCT 240 241 ATC TCC TGG CCC AGA ATT ATC CGS CST CAG CST CAGA GAS ALL TYL ATG Phe Ser BD
ATC TCC TCC ACC
241 ATC TCC TGG CCC AGA ATT ATG CCA GAT GGG AAG AAC ATC AAC CAA AAG GGT GTG GAT TTC 300 301 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AGA CTC GTT GAT GAG CTT TTC 300 101 TAC AAC AAC AGA CTC AGA GAG AAC AAC AAC AAC AAC AAC AAC AA
ASP Gly Lys Asn Ile Asn Gln Lys Gly TTC 300
301 TAC AAC AGA CTC GTT GAT GAG CTT TTG AAG AAT GAT ATC ATA CCA TTC GTA ACA CTC TAT 360
101 Tyr Asn Arg Leu Val Asp Glu Leu Leu Lys Asn Asp Ile Ile Pro Phe Val Thr Leu Tyr 120
361 CAC TOO COM TO THE DESCRIPTION OF THE PROPERTY AND THE LEU TOO.
121 HIS TOO ACT TAX CCC TAC GCA CTT TAT GAA AAA GCT CCA CTC CCA CTC TAX GAA AAA GCT CCA CTC CTC
361 CAC TGG GAC TTA CCC TAC GCA CTT TAT GAA AAA GGT GGA TGG CTT AAC CCA GAT ATA GCG 420
421 CTC TAT TTC ACA COS TO A 140
141 Leu Tyr Phe Arg Ala Tyr Ala Thr Phe Het Phe Asn Glu Leu Gly Asp Arg Val Lys His 160
Tyr Ala Thr Phe Het Phe Asn Glu Leu Gly Asn Are Lat 480
481 TGG ATT ACA CTG AAC GAA CCA TGG TGT TCT TCT TCT TCC GGT TAT TAC ACG GGA GAG CAT 540
161 TEP Ile The Leu Ash Glu Pro Tep Cys Ser Ser Phe Ser Gly Tyr Tyr The Gly Glu His 180
541 GCC CCC CCC CCC CCC CCC CCC CCC CCC CC
181 Ala Pro Civi W CAA AAT TTA CAA GAA GCG ATA ATC COO COO
541 CCC CCG GGT CAT CAA AAT TTA CAA GAA GCG ATA ATC GCG GCG CAC AAC CTC TTG AGG GAA .600 181 Ala Pro Gly His Gln Asn Leu Gln Glu Ala Ile Ile Ala Ala His Asn Leu Leu Arg Glu 200 601 CAT GGA CAT GCC GTC CAG GCG TCC ASC ASC ACC CTC TTG AGG GAA .600
OVI CAT GC: com and company and co
601 CAT GGA CAT GCC GTC CAG GCG TCC AGA GAA GAA GTA AAA GAT GGG GAA GTT GGC TTA ACC 660 201 His Gly His Ala Val Gln Ala Ser Arg Glu Glu Val Lys Asp Gly Glu Val Gly Leu Thr 220
Ala Ser Arg Glu Glu Val Lya Asp Gly Glu Val Cot TTA ACC 660
220 AAC GTT GTG ATG AAA ATA GAA CCG GGC CAT GGC
661 AAC GTT GTG ATG AAA ATA GAA CCG GGC GAT GCA AAA CCC GAA AGT TTC TTG GTC GCA AGT 720 721 CTT GTT GAT AAG TTC GTT AAR GGA AGT TCG GCA AGT 720
721 CIT GIT GIT ANG THE LOS PRO Glu Ser Phe Leu Val Ala Ser 240
721 CTT GTT GAT AAG TTC GTT AAT GCA TGG TCC CAT GAC CCT GTT GTT TTC GGA AAA TAT CCC 780 - 781 GAA GAA GCA GCA GTT GCA CTT TAR AGA CCA GTT GCA CTT TAR AGA CTT TAR
bys Phe Val Asn Ala Trp Ser His Asp Pro Val Val Trc GGA AAA TAT CCC 780
781 GAA CAA COA
261 Glu Glu Ala Val Ala Leu Tyr Thr Glu Lys Gly Leu Gln Val Leu Asp Ser Asp Het Asn 280
841 Am Ser Asp Ser Asp Mer Am 840
281 THE TEG ACT CCT ATA GAC TTC TTT CCT CTC ATA CAC TTC TTT CCT CTC ATA CAC TTC TT
841 ATT ATT TCG ACT CCT ATA GAC TTC TTT GGT GTG AAT TAT TAC ACA ACA ACA CTT GTT GTT 900 901 TTT GAT ATG AAC AAT CCT GTT GTT GTT GTT GTT GTT GAT ATG AAC ATG ATG AAC AAT CCT GTT GTT GTT GTT GAT ATG AAC AAT CCT GTT GTT GTT GTT GTT GAT ATG AAC AAT CCT GTT GTT GTT GTT GTT GTT GAT ATG AAC AAT CCT GTT GTT GTT GTT GTT GTT GAT ATG AAC AAT CCT GTT GTT GTT GTT GTT GTT GTT GTT GT
901 TIT Can be val val 300
901 TIT GAT ATG AAC AAT CCT CTT GGA TIT TCG TAT GTT CAG GGA GAC CTT CCC AAA ACG GAG 960 961 ATG GGA TGG GAA ATG TAG GGG GAG TGT CAG GGA GAC CTT CCC AAA ACG GAG 960 961 ATG GGA TGG GAA ATG TAG GGG GAG GAC CTT CCC AAA ACG GAG 960
ASA ASA Pro Leu Gly Phe Ser Tyr Val Gln Gly ASA Leu TAG GAG 960
961 ATC CCA magnetic
Het Gly Trp Glu Ile Tyr Pro Glo Con TAT TTT GAT ATG CTG GTC TAT CTG AAG CAN AGG
Het Gly Trp Glu Ile Tyr Pro Gln Gly Leu Phe Asp Het Leu Val Tyr Leu Lys Glu Arg 340
1021 TAT AAA CTA CCA CTT TAT ATC ACA GAG AAC GGG ATG GCT GGA CCT GAT AAA TTG GAA AAC 1080
By Leu Pro Leu Tyr Ile Thr Glu Asn Gly Het Ala Gly The GAA AAC 1080
1081 GGA AGA CTT CAT CAT CAT CAT CAT CAT CAT CAT CA
1081 GGA AGA GTT CAT GAT AAT TAC CGA ATT GAA TAT TTC GAA AAG CAC TTT GAA AAA GCA CTT 1140
361 Gly Arg Val His Asp Ash Tyr Arg Ile Glu Tyr Leu Glu Lys His Phe Glu Lys Ala Leu 380
1141 GAA GCA ATC ALT COL DISCOURT OF THE GIVE LYS Ala Leu 380
181 Glu Ala Ile Ash Ala Asp Val Ash law law and TTC ATT TGG TCT TTG ATG GAT AAC 1700
181 Glu Ala Ile Asn Ala Asp Val Asp Leu Lys Gly Tyr Phe Ile Trp Ser Leu Het Asp Asn 400
1201 TTC GAA TGG GCG TGC GGA TAC TCC AAA CGT TTC GGT ATA ATC TAC GTA GAT TAC AAT ACC 1260
THE GIR TEP AIM CYR GIY TYP SET LYS AND PIC GIV THE CAT ATC CTA CAT TAC AAT ACC 1260
126) CCA AAA ACC APA AND THE 420
1261 CCA AAA AGG ATA TRE AAA GAT TCA GCG ATG TGG TTG AAG GAA TTT CTA AAA TCT TAA 1317 421 Pro Lys Arg He Len Lys Asp Ser Ala Het Trp Leu Lys Glu Pho Lys Asp Tct TAA 1317
. Lys Ser End 419

Figure 2

STAPHYLOTHERMUS MARINUS GLYCOSIDASE - 12G COMPLETE GENE SEQUENCE 9/95

1 TTG ATA AGG TIT CCT GAT TAT TTC TTG TIT GGA AGA GGT AGA TCA TGG GAG GAG ATY (1 Met 11e Arg Phe Pro Asp Tyr Phe Leu Phe Gly Thr Ala Thr Ser Sec Him Re	
1 Met 11e Arg Phe Pro Asp Tyr Phe Leu Phe Gly Thr Ala Thr Ser Ser His Glo 11e G	
fy the Leu Phe Gly Thr Ala Thr Ser Ser His Glo Line	iAi: 60
OI GGT ANT ANC ATA TIT ANT GAT TGG TGG GAG TGG GAG	20
61 GGT AAT AAC ATA TIT AAT GAT TGG TGG GAG TGG GAG ACT AAA GGC AGG ATT AAG (FTG) 21 Gly Asn Asn Ile Phe Asn Asp Trp Trp Glu Trp Glu Thr Lys Gly Asn Lie	CA. 120
121 TCG CCT AAC CO	0.6 93/
121 TCG GGT AAG GCA TGT AAT CAT TGG GAA CTC TAT AAA GAA GAC ATA GAG CTT ATG GCT G 181 CTG GGA TAT AAT GCT TAT AGG TTA TAG GAA GAC ATA GAG CTT ATG GCT G 181 CTG GGA TAT AAT GCT TAT AGG TTA TAG TTA GAG TAT AGG TTA TAG GCT TAT AGG TTA TAG GTT TAT AGG TTA TAG TTAG	
als firp Glu Leu Tyr Lys Glu Asp Ile Glu Leu Her All o	AG 180
61 LOU CON TAT AAT GCT TAT AGG TTC TCC ATA GAG TCC AGA	10 60
181 CTG GGA TAT AAT GCT TAT AGG TTC TCC ATA GAG TGG AGT AGA ATA TTT CCC AGA AAA G 61 Leu Gly Tyr Asn Ala Tyr Arg Phe Ser Ile Glu Trp Ser Arg Ile Phe Pro Arg Lys A: 241 CAT ATA GAT TAT GAG TCC CTR AND	AT 240
241 CAT ATA CAT THE PIO ATG LYS A	5D RO
241 CAT ATA GAT TAT GAG TCG CTT AAT AAG TAT AAG GAA ATA GTT AAT CTA CTT AGA AAA TA Bl His Ile ASP Tyr Glu Ser Leu Asn Lys Tyr Lys Glu Ile Val Asp Lou Lou Asa AAA TA	
81 His Ile ASP TYR GAG TCG CTT AAT AAG TAT AAG GAA ATA GTT AAT CTA CTT AGA AAA TA 301 GGG ATA GAA CCT GTA ATC ACT CTA ATC ACT AGA ATA GTT AAT CTA CTT AGA AAA TA	300
301 GGG ATA GAA CCT GTA ATC ACT CTT CAC CAC TTC ACA AAC CCG CAA TGG TTT ATG AAA AT 101 Gly Ile Glu Pro Val Ile Thr Leu His His Phe Thr Asn Pro Gla Top The Ala AT	T 100
101 Gly Ile Glu Pro Val Ile Thr Leu His His Phe Thr Asn Pro Gln Trp Phe Het Lys Il	T 360
361 GGT GG1 TGC 100 100	0 120
121 Gly Gly TEP The Arg Glu Glu Asn Ile Lys Tyr Phe Ile Lys Tyr Val Glu Leu Ile Al. 121 TCC GAG ATA ANA GROUND AND AND AND AND AND AND AND AND AND A	T 435
421 TCC CAG and All Leu Ile All Toc CAG Tyr Val Glu Leu Ile All	T 420 B 140
421 TCC GAG ATA AAA GAC GTG AAA ATA TGG ATC ACT ATT AAT GAA CCA ATA ATA TAT GTT TTY 141 Ser Glu Ile Lys Asp Val Lys Ile Trp Ile Thr Ile Asp Glu Pro Ile Tat GTT TTY	
141 Ser Glu Ile Lys Asp Val Lys Ile Trp Ile Thr Ile Asn Glu Pro Ile Ile Tyr Val Let 481 CAA GGA TAT ATT TCC CCC Clare to the control of the c	480
481 CAA GGA TAT ATT TCC GGC GAA TGT CCA CCT CCC	160
481 CAA GGA TAT ATT TCC GGC GAA TGG CCA CCT GGA ATT AAA AAT TTA AAA ATA GGT GAT CAA 161 Glm Gly Tyr Ile Ser Gly Glu Trp Pro Pro Gly Ile Lys Asm Leu Lys Ile Ala Asp Glm 541 GTA ACT AAG AAT CTT TTD AAA AAT	540
541 GTA ACT ALC AND GENERAL ASP GIN	180
541 GTA ACT ANG ANT CIT ITA ANA GCA CAT ANT GNA GCC TAT ANT ATA CIT CAT ANA CAC GGT 181 Val Thr Lys Asn Leu Leu Lys Ala His Asn Glu Ala Tyr Asn Ile Leu His Lys His Gly 601 ATT GTA GGC ATA GCT ANA LAG CAC	
601 arm on the His Asn Glu Ala Tyr Asn Ile Leu His Lys His Gly	600 200
601 ATT GTA GGC ATA GCT ANA ANC ATG ATA GCA TTT ANA CCA GGA TCT ANT AGA GGA ANA GAC 201 Ile Val Gly Ile Ala Lys Asn Het Ile Ala Phe Lys Pro Gly Ser Asn Aga GAC	200.
201 Ile Val Gly Ile Ala Lys Asn Het Ile Ala Phe Lys Pro Gly Ser Asn Arg Gly Lys Asp 661 ATT AAT ATT TAT CAT Ala Company	
661 ATT AAT ATT TAT CAT ANA GTC GAT ANA GCA TTC AAC TGG GGA TTT CTC AAC GGA ATA TTA 221 Ile Asn Ile Tyr His Lys Val Asp Lys Ala Phe Asn Trp Gly Phe Loui And GGA ATA TTA	220
221 Ile Asn Ile Tyr His Lys Val Asp Lys Ala Phe Asn Trp Gly Phe Leu Asn Gly Ile Leu 721 AGG GGA GAA CTA GAA AGG GGA TAG GGA ATA TTA	720
721 AGG GGA CAA GTD DEL ASH GIY IIe Leu	240 .
721 AGG GGA GAA CTA GAA ACT CTC CGT GGA AAA TAC CGA GTT GAG CCC GGA AAT ATT GAT TTC 781 ATA GGC ATA AAC TAT THE TOTAL COLUMN TO THE Leu Arg Cly Lys Tyr Arg Val Clu Pro Cly Asn Ile Asp Phe	***
781 ATS CON AND THE ASP PRO	780 260
781 ATA GGC ATA AAC TAT TAT TCA TCA TAT ATT GTA AAA TAT ACT TGG AAT CCT TTT AAA CTA 261 Ile Gly Ile Asn Tyr Tyr Ser Ser Tyr Ile Val Lys Tyr The TTD Acc CCT TTT AAA CTA	
261 Ile Gly Ile Asn Tyr Tyr Ser Ser Tyr Ile Val Lys Tyr Thr Trp Asn Pro Phe Lys Leu	840
841 CAT ATT ANA GTC GAA CCA TTA GAT ACA GGT CTA TGG ACA ACT ATG GGT TAC TGC ATA TAT 281 His Ile Lys Val Glu Pro Leu Asp Thr Gly Leu Trp Thr Thr Her Club.	280
His He Lys Val Glu Pro Leu Asp Thr Gly Leu Ten Act ATG GCT TAC TGC ATA TAT	900
901 CCT AGA CCA AGA CCA	300
901 CCT AGA GGA ATA TAT GAA GTT GTA ATG AAA ACT CAT GAG AAA TAC GGC AAA GAA ATA ATC 301 Pro Arg Gly Ile Tyr Glu Val Val Het Lys Thr His Glu Lys Tyr Gly Lys Tyr Gl	
301 Pro Arg Gly 11e Tyr Glu Val Val Het Lys Thr His Glu Lys Tyr Gly Lys Glu 11e 11e	960 320
961 ATT ACA GAG AAC GGT GTT GCA GTA GAA AAT GAT GAA TTA AGG ATT TTA TCC ATT ATC AGG 121 Ile Thr Glu Asn Gly Val Ala Val Glu Asn Asp Glu Leu Arg Ile Leu Tra	320
121 Ile Thr Glu Asn Gly Val Ala Val Glu Asn Asp Glu Leu Arg Ile Leu Ser Ile Ile Arg 1021 CAC TTA CAA TAC TTA TAT Ass asp	1020
1021 CAC TTA CAA TAC TTA TAT AAA CCC ATT AND TO	340
1021 CAC TTA CAA TAC TTA TAT AAA GCC ATG AAT GAA GGA GCA AAG GTG AAA GGA TAT TTC TAC 141 His Leu Gln Tyr Leu Tyr Lys Ala Het Asn Glu Gly Ala Lys Val Lys Gly Tyr Phe Tyr 1081 TGG AGC TTC ATG GAT AND THE	1080
1081 TGG AGC TTC ATC COR	360
1081 TGG AGC TTC ATG GAT AAT TTT GAG TGG GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA 361 Trp Ser Phe Het Asp Asn Phe Glu Trp Asp Lys Gly Phe Asn Glo Arg Ttc GGA CTA GTA	
361 Trp Ser Phe Met Asp Asn Phe Glu Trp Asp Lys Cly Phe Asn Cln Arg Phe Gly Leu Val	1·140 380
181 GIU VAI ASP TYF LYS THE PHE GIU AFG LYS PEO AFG LYS SEE ALA TYF LYS TAT ACT CAA	
38) GIU VAI ASD TYF LYS THE PHE GIU ARG LYS PEO ARG LYS SEE ALA TYF VAI TYF SEE GID	1200
1201 ATA UCA COT ACC AND ACC AND ACC AND ACC AND ACC ACC ACC ACC ACC ACC ACC ACC ACC AC	400
401 Ile Ala Arg The Line ACT GAT GAA TAC CTA GAA AAA TAT COL	1260
1261 MA COLOR LYS ASD Leu	120
421 Glu End 422	
•••	

Figure 3

Thermococcis 9N2 Glydesidase -318/0 Complete gene sequence 9/95

ATG CTA CCA GAA GGC TIT CTC TGG GGC GTC TCC CAG TCC GGC TIT CAG TTC GAG Net Lau Pro Glu Gly Phe Leu Trp Gly val Ser Gla Ser Gly Phe Glo	
1 Met Lau Pro Clu Cly Phe Leu Trp Cly val Ser Cln Ser Cly Phe Glu Phe Glu 61 CAC AAG CTC ACC ACC ACC ACC ACC ACC ACC ACC ACC	ATC CCC
Tip Ciy Val Ser Cin Sex Ciy Phe Cin Phe Civ	ATC CCC 60
61 CAC AAG CTC AGG AAG AAC ATT GAT CCG AAC ACA GAC TGG TGG AAG TGG GTC AGG 21 Amp Lym Leu Ard Amp Ile Amp Pro Amp The Amp Tid Tid Lym	MEE GIA 30
21 Asp Lys Leu Ard Ard Ard Ard Ard Ard Ard GAC TOG TOG AND TOG COT	
And lie Amp Pro Am The Amp Tro Tro Live Tro	CY1 CCC 130
21 Amp Lym Leu Ard And Can The Amp Pro Am The Amp Trp Trp Lym Trp Val Arg 121 Tre Aac Ara Aag and Gaa cre ure and one cru cre gas can deg ara and (1 Phe Am Ile Lym Arg Glu Leu Val Ser Gly Amp Leu Pro Glu Glu Glu Glu	Asp Pro 40
41 Phe Ann the time and GAA CTC UTC ACC USC CAC CTU CCC CAC CTU	
ATE LYS AFR Glu Lau Val Ser Gly ARD Lau Dro Glu Che GGG ATA ARC	MC TAC 180
(1 Phe Amn 11e Lym Arg Glu Lmu Val Ser Gly Amp Leu Pro Glu Glu Gly 11e Amn 181 GAA CTT TAC GLG AAC CTT GLG	60 זער מפא
181 GAA CTT TAC GAG AAG GAT CAC CGC CTC GCC AGA GAC CTC GCT CTG AAC GTT TAC ; 61 Glu Leu Tyr Glu Lys Asp Him Arg Leu Ala Arg Asp Leu Gly Leu And Arg Arg Asp Leu Gly Leu And Arg Arg Arg Asp Leu Gly Leu And Arg	
THE DES TYP GIT LYS ASP MAN ASE LES A'S ASE	AGG ATT 240
61 Glu Leu Tyr Glu Lys Asp Him Arg Leu Ala Arg Asp Leu Gly Leu Ash Val Tyr) 241 GGA ATA GAG TGG AGG AGG	Ard Ile an
81 Gly 1le Glu Trp Ser Arg 1le Phe Pro Trp Pro Thr Trp Phe Val Glu Val	AB IIe 80
81 Gly Ile Glu Trp Ser Arg Ile Phe Pro Trp Pro Thr Trp Phe Val Glu Val Asp V	· · · ·
The Fro Thr Trp Phe Wal Glu Val and	300
301 CGG GAC AGC TAC GGA CTC GTG AAG GAC GTC AAA ATC GAT AAA GAC ACG CTC GAA G 101 Arg Asp Ser Tyr Gly Leu Val Lys Asp Val Lys I'e Asp Lys Am The Lys Asp	ar era 100
101 Arg Asp Ser Tyr Gly Leu Val Lys Asp Val Lys Ile Asp Lys Asp The Leu Glu G 161 GAC GAG ATA CCC AND COM COM COM CARD CARD CARD CARD CARD CARD CARD CARD	
ory and val Lys Asp val Lys I'm Asp Lys Am The Live GAR of	VC C3C 360
361 CAC CAC ATT COR ALL	lu Leu 120
121 ASP GLU Ile Ala ASE His Glu Glu Ile Ala Tyr Tyr Arg Arg Val Ile Glu His Li 421 GAG CTG GGC TTC LAG CTG	
The Ala Ash Ris Gln Glu Fie Ala Tot Tor Ira Ind And GAG CAC C	TC AGG 420
421 CAC COM COS TO CALL HIS LA	eu Ary 140
14) CIU CUE GOE TTO ANG CTC AND CTC AND CTC AND CTC AND CTC	
421 GAG CTG GGC TTC AAG GTC ATC GTG AAC CTG AAC CAC TTC ACG GTC CCC GTC TGC CTG 141 Glu Leu Gly Phe Lys Val Ile Val Asn Leu Asn His Phe Thr Leu Pro Leu Trp Le	FT CAC 480
And the The Leu Pro Leu Tro Leu Pro Le	ru His 160
481 GAT CCC ATA ATC CCG ACC GAG AAG GCT CTC ACC AAC GGT ACG ATT GCC TGC GTC GC 161 AFP PTO Ile Ile Ale Arg Glu Lys Ale Lou Thr Ann Gly Arg Tle Gly acc	
161 APP Pro The Hie Ala Arg Glu Lym Ala Lou The Ann Gly Arg The Gly Trp Val Gl	25 030 010
541 GAG ACC GTG GTG GAG TTC GGC AAG TAC GGG GGG TAC ATC GGG GAC GGA GGA GGA GGA GGA GGA GGA GG	
201 CTT CAT ATC TGG AGC ACC TTC AAC GAG CCG ATG GTC GTT GTG GAG CTC GGT TAC CTC GOT TAC CTC	
Val ASD Net TEP Ser The Phe Ash Slu Pro Het Val Val Val Glu Leu Gly Tyr Let	C GCG 660
PEO TYP Ser Gly Phe Pro Pro Gly Val Mer Act CEE GAS GCG GCA AAG CTG GCA ATT	
221 Pro Tyr Ser Gly Phe Pro Pro Gly Val Met Ast Pro Glu Ala Ala Lys Leu Ala II	CTC 720
721 AAC ATC ATS AND COO	CCC 720
721 AAC ATC ATS AND COO	CCC 720
721 AAC ATG ATA AAC GCC CAC GCA CTC CCC TAC AAG ATG ATA AAG AAG TTC GAC AGG GTA 241 Asn Het I'e Asn Ala His Ala Leu Ala Tyr Lys Het I'e Lys Ive The Aac AGG GTA	T CTC 720 Leu 240
721 AAC ATG ATA AAC GCC CAC GCA CTC CCC TAC AAG ATG ATA AAG AAG TTC GAC AGG GTA 241 Asn Het Ile Asn Ala His Ale Leu Ale Tyr Lys Het Ile Lys Lys Phe Asp Arg Val 781 GCC GGT AAG GTR GGG GTR	CCC 720 Leu 240 LAG 780 Lys 260
721 AAC ATG ATA AAC GCC CAC GCA CTC CCC TAC AAG ATG ATA AAG AAG TTC GAC AGG GTA 241 Asn Het Ile Asn Ala His Ale Leu Ale Tyr Lys Het Ile Lys Lys Phe Asp Arg Val 781 GCC GGT AAG GTR GGG GTR	CCC 720 Leu 240 LAG 780 Lys 260
721 AAC ATG ATA AAC GCC CAC GCA CTG CCC TAC AAG ATG ATA AAG AAG TTG GAC AGG GTZ 241 Asn Het Ile Asn Ala His Ale Leu Ale Tyr Lys Net Ile Lys Lys Phe Asp Arg Val 781 CCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC CCG ATA ATC TAC AAC AAC ATA GCC 251 Ale Asp Lys Asp Ser Arg Ser Glu Ale Glu Val Gly Ile Ile Tyr Acc Ata CCC	720 120 240 240 240 250 250 250 250 250 250 250 250 250 25
721 AAC ATG ATA AAC GCC CAC GCA CTC CCC TAC AAG ATG ATA AAG AAG TTC GAC AGG GTZ 241 Asn Het Ile Asn Ala His Ala Leu Ala Tyr Lys Het Ile Lys Lys Phe Asp Arg Val 781 CCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC GCG ATA ATC TAC AAC ATA GCC 251 Ala Asp Lys Asp Ser Arg Ser Glu Ala Glu Val Gly Ile Ile Tyr Asn Asn Ile Gly 841 GCC TAT TCA TAC GAG GAG	CCTC 720 8 Leu 240 1 RAG 780 1 Lys 260 1 CTT 840 1 Val 280
721 AAC ATG ATA AAC GCC CAC GCA CTC CCC TAC AAG ATG ATA AAG AAG TTC GAC AGG GTZ 241 Asn Het Ile Asn Ala His Ala Leu Ala Tyr Lys Het Ile Lys Lys Phe Asp Arg Val 781 CCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC GCG ATA ATC TAC AAC ATA GCC 251 Ala Asp Lys Asp Ser Arg Ser Glu Ala Glu Val Gly Ile Ile Tyr Asn Asn Ile Gly 841 GCC TAT TCA TAC GAG GAG	CCTC 720 8 Leu 240 1 RAG 780 1 Lys 260 1 CTT 840 1 Val 280
721 AAC ATG ATA AAC GCC CAC GCA CTC CCC TAC AAG ATG ATA AAG AAG TTC GAC AGG GTZ 241 Asn Het Ile Asn Ala His Ala Leu Ala Tyr Lys Het Ile Lys Lys Phe Asp Arg Val 781 CCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC GCG ATA ATC TAC AAC ATA GCC 251 Ala Asp Lys Asp Ser Arg Ser Glu Ala Glu Val Gly Ile Ile Tyr Asn Asn Ile Gly 841 GCC TAT TCA TAC GAG GAG	CCTC 720 8 Leu 240 1 RAG 780 1 Lys 260 1 CTT 840 1 Val 280
ALC ATG ATA AAC GCC CAC GCA CTG CCC TAC AAG ATG ATA AAG AAG TTG GAC AGG GTA 241 Asn Het Ile Asn Ala His Ala Leu Ala Tyr Lys Het Ila Lys Lys Phe Asp Arg Val 781 CCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC GGG ATA ATC TAC AAC AAC ATA CCC 251 Ala Asp Lys Asp Ser Arg Ser Glu Ala Glu Val Gly Ile Ila Tyr Asn Asn Ile Gly 241 GCC TAT CCA TAC GAC TCC AAC GAC CCA AAG GAC GTG AAA GCT GCA GAA AAC GAC AAC 281 Ala Tyr PTO Tyr Asp Ser Asn Asp Fro Lys Asp Val Lys Ala Ala Glu Lac	CCC 720 8 Leu 240 1 AAG 780 1 Lys 260 1 CCT 860 Val 280
ALC ATG ATA AAC GCC CAC GCA CTG CCC TAC ANG ATG ATA AAG AAG TITE GAC AGG GTA 241 Asn Het Ile Asn Ala His Ala Leu Ala Tyr Lys Net Ila Lys Lys Phe Asp Arg val 781 GCC GAT AAG GAT TCC GGC TCC GAG GCC GAG GTC GGG ATA ATC TAC AAC AAC ATA GGC 251 Ala Asp Lys Asp Ser Arg Ser Glu Ala Glu Val Gly Ile Ile Tyr Asn Asn Ile Gly 241 GCC TAT CCA TAC GAC TCC AAC GAC CCA AAG GAC GTG AAA GCT GCA GAA AAC GAC AAC 281 Ala Tyr 700 Tyr Asp Ser Asn Asp Fro Lys Asp Val Lys Ala Ala Glu Asn Asp Asn 901 TCC CAC AGG GCC GGG	CTC 720 8 Leu 240 1 AAG 780 1 Lys 260 2 CTT 840 2 Val 280 TAC 900 Tyr 300
ALC ATG ATA AAC GCC CAC GCA CTG CCC TAC ANG ATG ATA AAG AAG TITE GAC AGG GTA 241 Asn Het Ile Asn Ala His Ala Leu Ala Tyr Lys Net Ila Lys Lys Phe Asp Arg val 781 GCC GAT AAG GAT TCC GGC TCC GAG GCC GAG GTC GGG ATA ATC TAC AAC AAC ATA GGC 251 Ala Asp Lys Asp Ser Arg Ser Glu Ala Glu Val Gly Ile Ile Tyr Asn Asn Ile Gly 241 GCC TAT CCA TAC GAC TCC AAC GAC CCA AAG GAC GTG AAA GCT GCA GAA AAC GAC AAC 281 Ala Tyr 700 Tyr Asp Ser Asn Asp Fro Lys Asp Val Lys Ala Ala Glu Asn Asp Asn 901 TCC CAC AGG GCC GGG	CTC 720 8 Leu 240 1 AAG 780 1 Lys 260 2 CTT 840 2 Val 280 TAC 900 Tyr 300
ALC ATG ATA AAC GCC CAC GCA CTC CCC TAC AAG ATG ATA AAG ADG TTC GAC AGG GTZ ABH Het Ile ABH Ala His Ala Leu Ala Tyr Lys Met Ile Lys Lys Phe Aep Arg Val TRI CCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC GGG ATA ATC TAC AAC AAC ATA GCC Ala Abp Lys Abp Ser Arg Ser Glu Ala Glu Vai Gly Ile Ile Tyr Aem Aem Ile Gly Ali GCC TAT CCA TAC GAC TCC AAC GAC CCA AAG GAC GTG AAA GCT GCA GAA AAC GAC AAC BAL Tyr Pro Tyr Asp Ser Aem Aep Pro Lys Aep Val Lys Ale Ale Glu Aem Aep Aem TCC CAC AGC GGG CTC TTC TTC GAC GCA ATC CAC AAG GCC AAG CTC AAC ATC GAG TTC TO THE His Ser Gly Leu Phe Phe Aep Ala Ile His Lys Gly Lys Leu Aem Tle Gly Cor Tile Che Cac And CCC CAC CTC AAC ATC CAC TCC CAC CTC TCC TTC CTC GAC GCA ATC CAC AAG GCC AAG CTC AAC ATC GAG TTC TO THE His Ser Gly Leu Phe Phe Aep Ala Ile His Lys Gly Lys Leu Aem Tle Che Tile Che Tile Che Cac CTC AAC ATC CAC CTC AAC ATC CAC TCC CAC TTC TT	CCTC 720 8 Leu 240 1 AAG 780 1 Lys 260 1 CTT 840 1 Val 280 1 TAC 900 1 Tyr 300 CAC 960
ALC ATG ATA AAC GCC CAC GCA CTC CCC TAC ANG ATG ATA AAG ADG TITE GAC AGG GTA Ann Het Ile Ann Ala His Ala Leu Ala Tyr Lys Het Ile Lyn Lyn Phe And Arg Val The CCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC GGG ATA ATC TAC AAC AAC ATA GCC Ala And Lyn And Sat TCC CGC TCC GAG GCC GAG GTC GGG ATA ATC TAC AAC AAC ATA GCC Ala And Lyn And Sat TCC CAC TCC GAC GCC AAG GAC GTG AAA GCT GGA GAA AAC GAC AAC Ala Tyr Fro Tyr And Ser And And CAC CCA AAG GAC GTG AAA GCT GGA GAA AAC GAC AAC TCC CAC AGG GGG CTC TTC TTC GAC GCA ATC CAC AAG GCC AAG CTC AAC ATC GAG TTC TCC CAC AGG GGG CTC TTC TTC GAC GCA ATC CAC AAG GCC AAG CTC AAC ATC GAG TTC TGGT GAC ACC TCC TCC GTG GAG AND AND ALD ILE HIE LYN GLY LYN Leu Ann Ile Glu Phe	TAG 780 Lys 360 Lys 360 Val 280 TAC 900 Tyr 300 GAC 960 Asp 320
ALC ATG ATA AAC GCC CAC GCA CTC CCC TAC ANG ATG ATA AAG ADG TITE GAC AGG GTA Ann Het Ile Ann Ala His Ala Leu Ala Tyr Lys Het Ile Lyn Lyn Phe And Arg Val The CCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC GGG ATA ATC TAC AAC AAC ATA GCC Ala And Lyn And Sat TCC CGC TCC GAG GCC GAG GTC GGG ATA ATC TAC AAC AAC ATA GCC Ala And Lyn And Sat TCC CAC TCC GAC GCC AAG GAC GTG AAA GCT GGA GAA AAC GAC AAC Ala Tyr Fro Tyr And Ser And And CAC CCA AAG GAC GTG AAA GCT GGA GAA AAC GAC AAC TCC CAC AGG GGG CTC TTC TTC GAC GCA ATC CAC AAG GCC AAG CTC AAC ATC GAG TTC TCC CAC AGG GGG CTC TTC TTC GAC GCA ATC CAC AAG GCC AAG CTC AAC ATC GAG TTC TGGT GAC ACC TCC TCC GTG GAG AND AND ALD ILE HIE LYN GLY LYN Leu Ann Ile Glu Phe	TAG 780 Lys 360 Lys 360 Val 280 TAC 900 Tyr 300 GAC 960 Asp 320
AAC ATG ATA AAC GCC CAC GCA CTC CCC TAC ANG ATG ATA AAG AAG AAC AAC AAC AAC AAC AAC AAC AA	CCCC 720 Leu 240 LAG 780 Lys 260 CTT 840 Val 280 TAC 900 Tyr 300 GAC 960 Asp 320 TAC 1020
AAC ATG ATA AAC GCC CAC GCA CTG CCC TAC AAG ATG ATA AAG AAG TTG GAC AGG GTZ 241 Asn Het Ile Asn Ala His Ala Leu Ala Tyr Lys Net Ile Lys Lys Phe Aap Arg Val 781 CCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC GGG ATA ATC TAC AAC AAC AAC ATA GCC 251 Ala Asp Lys Asp Ser Arg Ser Glu Ala Glu Val Gly Ile Ile Tyr Asn Asn Ile Gly 841 GCC TAT CCA TAC GAC TCC AAC GAC CCA AAG GAC GTG AAA GCT GCA GAA AAC GAC AAC 281 Ala Tyr Pro Tyr Asp Ser Aen Asp Pro Lys Asp Val Lys Ala Ala Glu Asn Asp Asn 901 TTC CAC AGC GGG CTC TTC TTC GAC GCA ATC CAC AAG GCC AAG CTC AAC ATC GAG TTC 901 Phe His Ser Gly Leu Phe Pha Asp Ala Ile His Lys Gly Lys Leu Asn Ile Glu Phe 961 GGT GAG ACC TTC GTC AAA GTT CGG CAT CTC ACG GCG AAC GAC TGG ATA GGC GTT AAC 1021 TAC ACG AGI CAL GTG GTG AAA GTT CGG CAT CTC ACG GCG AAC GAC TGG ATA GGC GTT AAC	CCTC 720 8 Leu 240 1 RAG 780 1 Lys 260 2 CTT 840 2 Val 280 TAC 900 Tyr 300 TAC 960 Asp 320 TAC 1020 Tyr 340
AAC ATG ATA AAC GCC CAC GCA CTG CCC TAC AAG ATG ATA AAG AAG TTG GAC AGG GTZ 241 Asn Het Ile Asn Ala His Ala Leu Ala Tyr Lys Net Ile Lys Lys Phe Aap Arg Val 781 CCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC GGG ATA ATC TAC AAC AAC AAC ATA GCC 251 Ala Asp Lys Asp Ser Arg Ser Glu Ala Glu Val Gly Ile Ile Tyr Asn Asn Ile Gly 841 GCC TAT CCA TAC GAC TCC AAC GAC CCA AAG GAC GTG AAA GCT GCA GAA AAC GAC AAC 281 Ala Tyr Pro Tyr Asp Ser Aen Asp Pro Lys Asp Val Lys Ala Ala Glu Asn Asp Asn 901 TTC CAC AGC GGG CTC TTC TTC GAC GCA ATC CAC AAG GCC AAG CTC AAC ATC GAG TTC 901 Phe His Ser Gly Leu Phe Pha Asp Ala Ile His Lys Gly Lys Leu Asn Ile Glu Phe 961 GGT GAG ACC TTC GTC AAA GTT CGG CAT CTC ACG GCG AAC GAC TGG ATA GGC GTT AAC 1021 TAC ACG AGI CAL GTG GTG AAA GTT CGG CAT CTC ACG GCG AAC GAC TGG ATA GGC GTT AAC	CCTC 720 8 Leu 240 1 RAG 780 1 Lys 260 2 CTT 840 2 Val 280 TAC 900 Tyr 300 TAC 960 Asp 320 TAC 1020 Tyr 340
ALC ATG ATA AAC GCC CAC GCA CTC CCC TAC ANG ATG ATA AAG AAG AGG GEA AGG GEA 241 Asn Het I've Asn Ala His Ala Leu Ala Tyr Lyo Het I've Lyo Lyo Phe And Arg CCC 781 CCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC CGG ATA ATC TAC AAC AAC ATA CCC 251 Ala Asp Lyo Asp Ser Arg Ser Glu Ala Glu Val Gly I've I've Tyr Aen Aen I've Gly 241 GCC TAT CCA TAC GAC TCC AAC GAC CCA AAG GAC GTC AAA GCT GAA GAA AAC GAC AAC 281 Ala Tyr Pro Tyr Asp Ser Aen Aep Pro Lyo Aep Val Lyo Ala Ala Glu Aen Aep Aen 301 TTC CAC AGC GGG CTC TTC TTC GAC GCA ATC CAC AAG GCC AAG CTC AAC ATC GAC TTC 301 Phe His Ser Gly Leu Phe Pha Aep Ala I've His Lyo Gly Lyo Leu Aen I've Glu Phe 361 GGT GAC ACC TTC GTC AAA GTT CCG CAT CTC ACG CCC AAC GAC TCG ATA CGC GTT AAC 302 Gly Glu Thr Phe Val Lyo Val Arg His Leu Arg Gly Aen Asp Trp I've Gly Val Asn 1021 TAC ACG AGA GAA GTC GTC AGG TAT TCC GAG CCC AAC TTC CCG AGC ATA CCC CTG ATA 1031 Tyr Thr Arg Glu Val Val Arg Tyr Ser Glu Pro Lyo Phe Pro Ser I've Rea 1041 Tyr Thr Arg Glu Val Val Arg Tyr Ser Glu Pro Lyo Phe Pro Ser I've Rea 1051 Tyr Thr Arg Glu Val Val Arg Tyr Ser Glu Pro Lyo Phe Pro Ser I've Rea 1051 Tyr Thr Arg Glu Val Val Arg Tyr Ser Glu Pro Lyo Phe Pro Ser I've Rea 1051 Tyr Thr Arg Glu Val Val Arg Tyr Ser Glu Pro Lyo Phe Pro Ser I've Rea 1051 Tyr Thr Arg Glu Val Val Arg Tyr Ser Glu Pro Lyo Phe Pro Ser I've Rea 1051 Tyr Thr Arg Glu Val Val Arg Tyr Ser Glu Pro Lyo Phe Pro Ser I've Rea 1052 Tyr	TAC 1080
ALC ATG ATA AAC GCC CAC GCA CTC CCC TAC ANG ATG ATA AAG AAG AGG GEA AGG GEA 241 Asn Het I've Asn Ala His Ala Leu Ala Tyr Lyo Het I've Lyo Lyo Phe And Arg CCC 781 CCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC CGG ATA ATC TAC AAC AAC ATA CCC 251 Ala Asp Lyo Asp Ser Arg Ser Glu Ala Glu Val Gly I've I've Tyr Aen Aen I've Gly 241 GCC TAT CCA TAC GAC TCC AAC GAC CCA AAG GAC GTC AAA GCT GAA GAA AAC GAC AAC 281 Ala Tyr Pro Tyr Asp Ser Aen Aep Pro Lyo Aep Val Lyo Ala Ala Glu Aen Aep Aen 301 TTC CAC AGC GGG CTC TTC TTC GAC GCA ATC CAC AAG GCC AAG CTC AAC ATC GAC TTC 301 Phe His Ser Gly Leu Phe Pha Aep Ala I've His Lyo Gly Lyo Leu Aen I've Glu Phe 361 GGT GAC ACC TTC GTC AAA GTT CCG CAT CTC ACG CCC AAC GAC TCG ATA CGC GTT AAC 302 Gly Glu Thr Phe Val Lyo Val Arg His Leu Arg Gly Aen Asp Trp I've Gly Val Asn 1021 TAC ACG AGA GAA GTC GTC AGG TAT TCC GAG CCC AAC TTC CCG AGC ATA CCC CTG ATA 1031 Tyr Thr Arg Glu Val Val Arg Tyr Ser Glu Pro Lyo Phe Pro Ser I've Rea 1041 Tyr Thr Arg Glu Val Val Arg Tyr Ser Glu Pro Lyo Phe Pro Ser I've Rea 1051 Tyr Thr Arg Glu Val Val Arg Tyr Ser Glu Pro Lyo Phe Pro Ser I've Rea 1051 Tyr Thr Arg Glu Val Val Arg Tyr Ser Glu Pro Lyo Phe Pro Ser I've Rea 1051 Tyr Thr Arg Glu Val Val Arg Tyr Ser Glu Pro Lyo Phe Pro Ser I've Rea 1051 Tyr Thr Arg Glu Val Val Arg Tyr Ser Glu Pro Lyo Phe Pro Ser I've Rea 1051 Tyr Thr Arg Glu Val Val Arg Tyr Ser Glu Pro Lyo Phe Pro Ser I've Rea 1052 Tyr	TAC 1080
AAC ATG ATA AAC GCC CAC GCA CTC CCC TAC ANG ATG ATA AAG AAG ACG ACG GTA AAG AAC ATA AAC AAA AAC ATA AAC CCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC CGG ATA AAC TAC AAC AAA AAC ATA CCC CGT AAA AAC TAC CAC CAC AAG CCC GAT CCC AAC CCC AAC CCC AAG CCC AAC CCC ACC ACC CCC ACC ACC CCC ACC AC	CCTC 720 Leu 240 LAG 780 Lys 260 CTT 840 Val 280 TMC 900 TYT 300 CAC 960 ASD 320 TAC 1020 TYT 340 TCC 1080 Ser 360
AAC ATG ATA AAC GCC CAC GCA CTC CCC TAC ANG ATG ATA AAG AAG ACG ACG GTA AAG AAC ATA AAC AAA AAC ATA AAC CCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC CGG ATA AAC TAC AAC AAA AAC ATA CCC CGT AAA AAC TAC CAC CAC AAG CCC GAT CCC AAC CCC AAC CCC AAG CCC AAC CCC ACC ACC CCC ACC ACC CCC ACC AC	CCTC 720 Leu 240 LAG 780 Lys 260 CTT 840 Val 280 TMC 900 TYT 300 CAC 960 ASD 320 TAC 1020 TYT 340 TCC 1080 Ser 360
AAC ATG ATA AAC GCC CAC GCA CTG CCC TAC ANG ATG ATA AAG AAG AAC AAC AAG GCT AAC AAC AAC AAA AAC AA	CCTC 720 8 Leu 240 1 AAG 780 1 Lys 260 1 CTT 840 1 Val 280 1 TAC 900 1 TYT 300 1 CAC 960 1 ASP 320 1 TAC 1020 1 TAC 1080
AAC ATG ATA AAC GCC CAC GCA CTC CCC TAC ANG ATG ATA AAG AAG AGG GEN AAG GAL AAG ATA AAC AAA AAC AA	CCTC 720 8 Leu 240 1 AAG 780 1 Lys 260 1 CTT 840 1 Val 280 1 TAC 900 Tyr 300 TAC 1020 TYC 1080 Ser 360 CGA 1140 Cly 380
AAC ATG ATA AAC GCC CAC GCA CTC CCC TAC ANG ATG ATA AAG AAG AGG GEN AAG GAL AAG ATA AAC AAA AAC AA	CCTC 720 8 Leu 240 1 AAG 780 1 Lys 260 1 CTT 840 1 Val 280 1 TAC 900 Tyr 300 TAC 1020 TYC 1080 Ser 360 CGA 1140 Cly 380
AAC ATG ATA AAC GCC CAC GCA CTG CCC TAC ANG ATG ATA AAG AAG ACG AAC AAC AAC AAC AAC AAC AA	CTC 720 Let 240 LAG 780 Lys 260 CTT 840 Val 280 TAC 900 TYT 300 CAC 960 Asp 320 TAC 1020 TYC 1080 Ser 360 CGA 1140 Cly 380 AGA 1200
AAC ATG ATA AAC GCC CAC GCA CTG CCC TAC ANG ATG ATA AAG ANG TITE GAC AGG GCA AAG HE IIe ASH AIE HIS AIE LEU AIE TYT LYS HET IIE LYS LYS Phe AAG ATG GCA AAG GCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC GCG ATA ACC TAC AAC AAC ATA CCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC GCG ATA ACC TAC AAC AAC ATA CCC AAA AAC ACA TAC CAC AAG GCC TAT CCA TAC GAC TCC AAC GAC CCA AAG GAC GCG AAA GCC GAA AAC GAC AAC A	TAC 1020 TYT 340 TAC 1020 TYT 300 TAC 1020 TYT 340 TAC 1020 TYT 340 TAC 1020 TYT 340 TAC 1080 Ser 360 CGA 1140 Cly 380 AGA 1200 AGA 1200 AGA 1200 AGA 1200 AGA 1200 AGA 1200
AAC ATG ATA AAC GCC CAC GCA CTG CCC TAC ANG ATG ATA AAG ANG TITE GAC AGG GCA AAG HE IIe ASH AIE HIS AIE LEU AIE TYT LYS HET IIE LYS LYS Phe AAG ATG GCA AAG GCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC GCG ATA ACC TAC AAC AAC ATA CCC CAT AAG GAT TCC CGC TCC GAG GCC GAG GTC GCG ATA ACC TAC AAC AAC ATA CCC AAA AAC ACA TAC CAC AAG GCC TAT CCA TAC GAC TCC AAC GAC CCA AAG GAC GCG AAA GCC GAA AAC GAC AAC A	TAC 1020 TYT 340 TAC 1020 TYT 300 TAC 1020 TYT 340 TAC 1020 TYT 340 TAC 1020 TYT 340 TAC 1080 Ser 360 CGA 1140 Cly 380 AGA 1200 AGA 1200 AGA 1200 AGA 1200 AGA 1200 AGA 1200
AME ATG ATA AMG GCC CAC GCA CTC CCC TAC ANG ATG ATA AMG AND TITE GAC AGG GCA AND Het Ile Ann Ala His Ale Leu ale Pyr Lyn Het Ile Lyn Lyn Phe amp Arg Val GCC CAT AMG GAT TCC GGC TCC GAG GCC GAG GTC GGG ATA ATC TAC AAC AAC AAC ATA GCC Ala Amp Lyn Amp Ser Arg Ser Glu Ala Glu Val Gly Ile Ile Tyr Amm Amn Ile Gly Ala Amp Lyn Amp Ser Arg Ser Glu Ala Glu Val Gly Ile Ile Tyr Amm Amn Ile Gly Ala Tyr Pro Tyr Amp Ser Amn Amp Pro Lyn Amp Val Lyn Ala Ala Glu Amm Amp Amn JOI TTC CAC AGC GGG CTC TTC TTC GAC GCA ATC CAC AAG GCC AAG CTC AAC ATC GAG TTC JOI Phe His Ser Gly Leu Phe Phe Amp Ala Ile His Lyn Gly Lyn Leu Amn Ile Glu Phe JOI GCT GAC ACC TTC GTC AAA GTT CCC CAT CTC AGG GGG AAC GAC TCG ATA GGC GTT AAC JOI GT GAC ACC TTC GTC AAA GTT CCC CAT CTC AGG GGG AAC GAC TCG ATA GGC GTT AAC JOI TAC ACG AGA GAA GTC GTC AGG TAT TCC GAG CCC AAC TTC CCG AGC ATA CCC CTG ATA LOI TAC ACG AGA GAA GTC GTC AGG TAT TCC GAG CCC AAC TTC CCG AGC ATA CCC CTG ATA JOI TAC ACG AGA GAA GTC GTC AGG TAT TCC GAG CCC AAC TTC CCG AGC ATA CCC CTG ATA JOI TAC ACG GGA GTT CAC AAC TAC GGC TAC GGC TGC AGG CCC AGT TCT TCC GGC GAC JOI TAC ACG GGA GTT CAC AAC TAC GGC TAC GGC TGC AGG CCC AGT TCT TCC GGC GAC JOI TAC ACG GGA GTT CAC AAC TAC GGC TAC GGC TGC AGG CCC AGT TCT TCC GGC GAC JOI TAC ACG GGA GTT CAC AAC TAC GGC TAC GGC TGC AGG CCC AGT TCT TCC GGC GAC JOI TAC ACG GGA GTT CAC AAC TAC GGC TAC GGC TGC AGG CCC AGT TCT TCC GGC GAC JOI TAC ACG GGA GTT CAC AAC TAC GGC TAC GGC TGC AGG CCC AGG GGG ATC TAC GAC TCC GAC JOI TAC ACG GGA GTT CAC AAC TAC GGC TGG GGG ATC TAC GAC TCC GAC JOI TAC ACG GGA GTC GAC ATC GGC TGG GAG ATC TAT CCC GAG GGG ATC TAC GAC TCC GAC JOI TAC ACG GGC AAC AAC AAA TAC GGC TGG GGG ATC TAC GGC GAC ACG GGA GCC AAC AAC AAA TAC GGC CTC CCC GTT TAC GTC ACC GAA AAC GGA ATA GCC GAT TCA ACG GCC AAC AAA AAC GGC CTC CCC GTT TAC GTC ACC GAA AAC GGA ATA GCC GAT TCA ACG GGA GCC AAC AAC AAA TAC GGC CTC CCC GTT TAC GTC ACC GAA AAC GGA ATC GCC GAT TCA ACG GTG AAC AAC AAA TAC GGC CTC CCC GTT TAC GTC ACC GAA AAC GGA ATA GCC GAT TCA ACG GTG AAC AAC AAC AAA TAC GGC CT	TAC 1020 TYC 1080 TYC
AME ATG ATA AMG GCC CAC GCA CTC CCC TAC ANG ATG ATA AMG AND TITC GAC AGG GCA ASS Het Ile Ass Ala His Ale Leu Ale Pyr Lys Het Ile Lys Lys Phe amp Arg Val 781 CCC CAT AMG GAT TCC GGC TCC GAG GCC GAG GTC GGG ATA ATC TAC AAC AAC AAC ACA 252 Ala Asp Lys Asp Ser Arg Ser Glu Ala Glu Val Gly Ile Ile Tyr Amm Ams Ile Gly 841 GCC TAT CCA TAC GAC TCC AAC GAC CCA AAG CAC GTG AAA GCT GGA GAA AAC CAC AAC 281 Ala Tyr Pro Tyr Amp Ser Amm Amp Pro Lys Amp Val Lys Ala Ala Glu Amm Amp Amm 301 TTC CAC AGC GGG GTC TTC TTC GAC GCA ATC CAC AAG GCC AAG CTC AAC ATC GAC TTC 301 Phe His Ser Gly Leu Phe Pha Amp Ala Ile His Lym Gly Lym Leu Amm Ile Glu Phe 361 GGT GAC ACC TTC GTC AAA GTT CGC CAT CTC AGG CGC AAC GAC TGC ATA GGC GTT AAC 302 GTT GAC ACC TTC GTC AAA GTT CGC CAT CTC AGG CGC AAC GAC TGC ATA GGC GTT AAC 303 TAC ACG AGA GAC GTC GTC AGA TAT TCC GAG CCC AAC TTC CGG ATA GGC GTT AAC 304 TAC ACG AGA GAA GTC GTC AGA TAT TCC GAG CCC AAC TTC CGG AGC ATA CCC CTG ATA 305 TTC CGC GGA GTT CAC AAC TAC GGC TAC GGC CCC AAC TTC CGG AGC ATA CCC CTG ATA 306 TTC CGC GGA GTT CAC AAC TAC GGC TAC GGC TGC AGG CCC AAC TTC TCC GCC GAC 307 TAC ACG CCC GTA AGC GAC ATC GGC TAC GGC TGC AGG CCC AGC TCT TCC GCC GAC 308 TTC CGC GGA GTT CAC AAC TAC GGC TAC GGC TGC AGG CCC AGT TCT TCC GCC GAC 309 TTC CGC GGA GTT CAC AAC TAC GGC TGC GGC TGC AGG CCC ACC TTC TCC GCC GAC 3101 TAC ACG CCC GTA AGC GAC ATC GGC TGG GAG ATC TAT CCC GAG GGG ATC TAC GAC TCC ATA 3101 GAG GCC AAC AAC AAC GGC TGG GAG ATC TAT CCC GAC GGG ATC TAC GAC TCC ATA 3101 GAG GCC AAC AAA TAC GGC TGC GGG GTT TAC GTC ACC GAA AAC GGA ATA GCC CAT TCA 401 Glu Ala Amm Lym Tyt Gly Val Pro Val Tyr Val Thr Glu Amn Cly Ile Ala Amp Ser Ile 3161 GAC ACC CTG CGG CTG TAC GCC TTC TAC GTC ACC GAA AAC GGA ATA GCC CAT TCA 401 Glu Ala Amm Lym Tyt Gly Val Pro Val Tyr Val Thr Glu Amn Cly Ile Ala Amp Ser Ile 3161 GAC CCC GTG CGC CTT TAC GTC TAC GAC ACC GAC ACC ACC ACC ACC ACC ACC A	TAG 780 Lys 260 Lys 260 Lys 260 Lys 260 The 900 Tyr 300 GAC 960 Asp 320 TAC 1020 Tyr 340 TCC 1080 Ser 360 CGA 1140 Cly 380 AGA 1200 Arg 400 Arg 400 Arg 420
AME ATG ATA AMG GCC CAC GCA CTC CCC TAC ANG ATG ATA AMG AND TITC GAC AGG GCA ASS Het Ile Ass Ala His Ale Leu Ale Pyr Lys Het Ile Lys Lys Phe amp Arg Val 781 CCC CAT AMG GAT TCC GGC TCC GAG GCC GAG GTC GGG ATA ATC TAC AAC AAC AAC ACA 252 Ala Asp Lys Asp Ser Arg Ser Glu Ala Glu Val Gly Ile Ile Tyr Amm Ams Ile Gly 841 GCC TAT CCA TAC GAC TCC AAC GAC CCA AAG CAC GTG AAA GCT GGA GAA AAC CAC AAC 281 Ala Tyr Pro Tyr Amp Ser Amm Amp Pro Lys Amp Val Lys Ala Ala Glu Amm Amp Amm 301 TTC CAC AGC GGG GTC TTC TTC GAC GCA ATC CAC AAG GCC AAG CTC AAC ATC GAC TTC 301 Phe His Ser Gly Leu Phe Pha Amp Ala Ile His Lym Gly Lym Leu Amm Ile Glu Phe 361 GGT GAC ACC TTC GTC AAA GTT CGC CAT CTC AGG CGC AAC GAC TGC ATA GGC GTT AAC 302 GTT GAC ACC TTC GTC AAA GTT CGC CAT CTC AGG CGC AAC GAC TGC ATA GGC GTT AAC 303 TAC ACG AGA GAC GTC GTC AGA TAT TCC GAG CCC AAC TTC CGG ATA GGC GTT AAC 304 TAC ACG AGA GAA GTC GTC AGA TAT TCC GAG CCC AAC TTC CGG AGC ATA CCC CTG ATA 305 TTC CGC GGA GTT CAC AAC TAC GGC TAC GGC CCC AAC TTC CGG AGC ATA CCC CTG ATA 306 TTC CGC GGA GTT CAC AAC TAC GGC TAC GGC TGC AGG CCC AAC TTC TCC GCC GAC 307 TAC ACG CCC GTA AGC GAC ATC GGC TAC GGC TGC AGG CCC AGC TCT TCC GCC GAC 308 TTC CGC GGA GTT CAC AAC TAC GGC TAC GGC TGC AGG CCC AGT TCT TCC GCC GAC 309 TTC CGC GGA GTT CAC AAC TAC GGC TGC GGC TGC AGG CCC ACC TTC TCC GCC GAC 3101 TAC ACG CCC GTA AGC GAC ATC GGC TGG GAG ATC TAT CCC GAG GGG ATC TAC GAC TCC ATA 3101 GAG GCC AAC AAC AAC GGC TGG GAG ATC TAT CCC GAC GGG ATC TAC GAC TCC ATA 3101 GAG GCC AAC AAA TAC GGC TGC GGG GTT TAC GTC ACC GAA AAC GGA ATA GCC CAT TCA 401 Glu Ala Amm Lym Tyt Gly Val Pro Val Tyr Val Thr Glu Amn Cly Ile Ala Amp Ser Ile 3161 GAC ACC CTG CGG CTG TAC GCC TTC TAC GTC ACC GAA AAC GGA ATA GCC CAT TCA 401 Glu Ala Amm Lym Tyt Gly Val Pro Val Tyr Val Thr Glu Amn Cly Ile Ala Amp Ser Ile 3161 GAC CCC GTG CGC CTT TAC GTC TAC GAC ACC GAC ACC ACC ACC ACC ACC ACC A	TAG 780 Lys 260 Lys 260 Lys 260 Lys 260 The 900 Tyr 300 GAC 960 Asp 320 TAC 1020 Tyr 340 TCC 1080 Ser 360 CGA 1140 Cly 380 AGA 1200 Arg 400 Arg 400 Arg 420
AME ATG ATA AMG GCC CAG GCA CTC GCC TAC ANG ATG ATA AMG AND TITE GAC AGG GCA ASS HET I'VE ASS ALLE HIS ALLEU ALLE TAY LYS HET I'VE LYS Phe AMS AND ATG GAC THE CCC CAT ANG GAT TCC GGC TCC GAG GCC GAG GTC GGG ATA ATC TAC AAC AAC AAC ACA ASS LYS AMS SET ANG SET GLU ALLE GLU VAL GLY I'VE LILE TYT AMS AMS I'VE GLY ALLA ASS LYS AMS SET AND AMS FOR CCA AAG GAC GTG AAA GCT GCA GAA AAC GAC AAC ALL TAY PTO TYT AMS SET AMS AMS FOO LYS AMS VAL LYS ALLE ALLE GLU AMS AMS BOLL TCC CAC AGG GGG GTC TTC TTC GAC GCA ATC CAC AAG GCC AAG CTC AAC ATC GAC ATC GAC JOLL PHE RIS SET GLY LEU PHE PHE AMS ALLE I'VE GLY LYS LEU AMS I'VE GLY LAS AMS I'VE GLY LAS AMS I'VE GLY LYS LEU AMS I'VE GLY LYS LEU AMS I'VE GLY LAS AMS GCT GAC ACC TTC GTC AAA GTT CCC CAT CTC AGG CGC AAC GAC TCG ATA GGC GTT AAC JOLL TAC ACG AGA GAA GTC GTC AAA GTT CCC CAT CTC AGG CGC AAC GAC TCG ATA GGC GTT AAC LOUI TAC ACG AGA GAA GTC GTC AGA TAT TCC GAG CCC AAC TTC CCG AGC ATA CCC CTG ATA LOUI TAC ACG AGA GAA GTC GTC AGG TAT TCC GAG CCC AAC TTC CCG AGC ATA CCC CTG ATA LOUI TAC ACG GGA GTT CAC AAC TAC GGC TAC GGC CCC AAC TTC CCG AGC ATA CCC CTG ATA LOUI TTC CCG GGA GTT CAC AAC TAC GGC TAC GGC TGC AGG CCC AGT TCT TCC GCC GAC ACG CCC GTA AGC GAC ATC GGC TGG GAG ATC TAC CCG GAG GGG ATC TAC GAC TCC GAC ACG CCC GTA AGC GAC ATC GGC TGG GAG ATC TAT CCC GAG GGG ATC TAC GAC TCC ATA ACG CCC GTA AGC GAC ATC GGC TGG GAG ATC TAT CCC GAG GGG ATC TAC GAC TCC ATA ACG FTO VAL SCT AMS I'VE TYP GLY TYP FTO CLU GLY I'VE AMS SET I'VE GAG GCC AAC AAA TAC GGC CTC CCC GTT TAC GTC ACC GAA AAC GGA ATA GCC CAT TCA LOUI GAL ALLA AMS LYS TYT GLY VAL PYO VAL TYP VAL THE GLU AMS CJY I'VE AMS SET I'VE COL GAC CCC CTC CCC CTT TAC GTC ACC GAA AAC GGA ATA GCC CAT TCA LOUI GAL ALLA AMS LYS TYT GLY VAL PYO VAL TYP VAL THE GLU AMS CJY I'VE AMS SET I'VE COL GAC CCC CTC CCC CTT TAC GTC ACC GAA AAC GGA ATA GCC CAT TCA LOUI GAL ALLA AMS LYS TYT GLY VAL PYO VAL TYP VAL THE GLU AMS CJY I'VE ALE AMS SET I'VE COL GAC CCC CTC CCC CTT TAC GTC ACC GAA AAC GGA ATA GCC CAT TCA LOUI GAL ACC CTG CGC CTC TAC ACC GTT TAC	TAG 780 Lys 260 Lys 260 Lys 260 Lys 260 The 900 Tyr 300 GAC 960 Asp 320 TAC 1020 Tyr 340 TCC 1080 Ser 360 CGA 1140 Cly 380 AGA 1200 Arg 400 Arg 400 Arg 420

Figure 4a

1321 441 .381	Ala	61	TAC	GAC	GT	. ACC	Coc	: TAC	: כדי	TAC	. TCG	CCC			•					GCC	
		,	. , , ,	VEL	Va 1	YEC	GI)	זע: י	Len	TV	Tm	41.		YCC	CYC	ANC	TAC	عدة	TCC	CCC	1100
. 381	CTC	CCT	*										_		410		LAX	Clu	113	Ala	450
461	Leu	Giv	Pho	A00	ATC	YCC	220	600	CTC	TAT	222	CTC:	·						•	****	460
	_	,		vza	Met	YLD	Phe	Gly	Lui	Tyr	Lvs	Val	4	CTC.	ATA	YCC	AAG	CAC	AGA	Ala	1440
1441	CCC	· ccc	*													4nr	ry.	G) y	Ara	The	480
481	Pro	Ara	Chi	Chi	YEC	CTA	WC	CTT	TAT	AGC	CCC	177							•		
		,	414	CIR	2et	.Val	Lye	Val	tyr	Are	Giv	Tim	V- 1	CAG	wc	YYC	CCY	OIC	ACC	MC	1500
150;	CAA	ATC	~										141	OTA	YID	λan	CIA	Val	Ser	Lvs	300
501	Clu	11.	450	CAG	770	21.0	CCY	CĻI	CFG	TGA	15	30								-, -	300
	_		<i>-</i>	CIU	-} . =	Pho	Cly	Leu	Gly	End	31										

Figure 4b(Continued)

	1	ATG Mei	GAA			rc G	AT C	ÄA A	л (דכ	זרד	("AG	TT	٠ ٨	.	ACA	GAG	GAZ	١ ٨٨	ះ ភា	G A	v: ~	**		
				·	e Ihr					20	Scr	Cin	l.ev	Th	ır 7	Thr	Clu	Glu	l.yx			_		CTT Val	60 20
		CTC Val	GGG		T GO		TT (CA C	GA C		TTT Phe	GUG	AA(Asn	C CC	A (ΑT			េកប	ט טכ	C C	iT G		CT	120
ı	21	GGA	GAA	AC	, V CY	T C						-				lis	Ser	Arg		Λla	Gly	, A1	, د	Ma	40
	41	Gly	Clu	Thr	His	Pr	n V	al Pr	10 A	irg 1	בון בפט	GGA	ATT	Pre		ICG Na	TTT Phe	CTC Vai	CTO Lev	VP CC				cc	130
	81	GCA	GGA	СТО	AG	A A7	ΓΑ . Α	AT C	CC A	CA A	NGG	GAA	AAC	: GA	J G	AA	AAC	A CT	TAC		_			ro	60
			•		•				.,	nr /	vrg.	Ciu	Asn.	Ası	, 0	iu	Asn	Thr	Tyr	TA(,-			iCA .la	240 80
24	41 ' 31 1	TT The	CCC Pho	CTT	GA. Glu	A AT	M J	TG C	כ כ	כד ז	CT	ACC	TGG	*			GAC	сп	CTG	GA,	A GA	А СТ	6 6	GA	300
30) I		CCC									Tur	Τm	معہ	-	-	Asp	Lev	Lev	Glu	Glu	Val		ly	100
10	i	ys .	Ala	Mei	Gly	A GA Glu	A G/	V.	T A	GG G	AA lu	TAC Fyr	GGT Gly	Val			GTG Val	CTT Lev		GCA			G A	TG	360
36	1 1	AC,	ATT -	CAC	: AG	A AÁ Asn	c co					•	•			•			Lev	Ala	Pro	Ala	М	C1	120
12	.1 A	.sn	le	His	Arg	Asn	Pro	Le	Су	s G	ly /	\rg	Asn	Phe	. G)			TAC Tyr	TCA Ser	GA.A Glu	GAT Asp	CC Pro	ר כי	rc	420 140
42 14	_	TT 7	רככ ומ	GCT Gly	GAA	AT(G GC	T TC	A GC	C T	п с	πc .	MG	GGA	, G	n (244	TCT	CAA	GGG					
48				•			• • • •	~	~"	r	iE 4	11	Lys	Gly	Va	1 (Sin	Ser	Gin	Giy	Val	Gly	A)	C	480 160
161	-	ys II		AAA Lys		Phe	Val	C GC	0 ۸۸ تعام	.C A.	AC C	AG (GAA	ACC Thr		C A	CG	ATG		CTG	GAC	ACC	; AT	c	540
541	G	rg T	cc	GAG	CGA	GCC	: CT											Ma	Val	Val	Аæр	Thr	lle		180
181								AL S	CIB	ile	17	yτ L	.eu	Lys	Gly	P	he (GAA Glu	lle	GCT Ala	GTC Val	AAC Lys	Ly:		600 200
601 201	G	: A A	GA (200	TGG	ACC	GT(ATO	A G	C 60	т т.	AC A	мc	***	СТ	G A	AT (GGA	AAA	TAC	тст	TCA	CA.		660
661					-					Д			rasu.	Lys	Leu		ru (Sty	Lys	Tyr	Cys	Ser	Gin		220
221	٠.٠	n Gi	ל טו	TP	Lev	TTG Leu	Lys	i AAC Lys	Val	CT Les	C A	GG G	AA N	GAA Glu	TG(G G			GGC	CCT	TTC	cτc	AT		720
721 241	AG	c c	AC T	.cc	TAC	GCG Ala	GGA	GAC		: cc		-						jie.	Cly	Cly	Phe	Vel	Met	•	240
					•		- ,		74311	FID	•	, ,	1U ·	Gin	Leu	Ly	'S A	da d	GGA Gly	AAC	GAT Asp	ATG Mci	AT(780 260
78 I 26 I	AT: Mei	G CC	.T G	GG /	AAA .vs	GCG Ala	TAT	CAG Gln	CTC	AA	- AC	A G	м,	AGA	AGA	GA	T G		ATA	GAA	GAA	ATC	ATC		840
841							•			7011	* 411	· O	u /	^rg	Arg	As;	P G		le	Clu	Glu	He	Met		280
23 1	Glu	Ala	L	eu L	.ys	GAG Glu	Cly	Lys	Leu	Ser	GA Glu	G G/ Gi	, O	JII ∕ai	CTC Leu	GA Asp		AG 7	GT Sys	GTG Val	AGA	MC	ATT		200
30 I 106	CTC	Lys	A G	יד כ	11	GTG.	AAC	GCG	сст	TCC	TI	- AA	.A C	CG	TAC		G T				Arg	Asn	lie		100
961		_,	•		LU	A 81	Asn	Ata	Pro	Ser	Phe	Ly:	s (ily	Tyr	Arg				AAC Asn	AAG Lys	CCG	GAT Asp		60 20
321	Leu	Clu	A TC	т С. Г Н	AC is	GCG Ala	GAA Giv	CTC Val	GCC	TAC Tyr	GA/ Glu	GC					G G	ज द	TT.	стс	стт	כדד	GAG	10	20
1021	AAC	AÁC	: GC	п с	π,	(11	ccc			•		Ala		•	Ala	Glu	G1	y v	al	Vai	Lev	Lev	Clu	34	40
341	Asn	Asn	Giy	/ V:	1			rat /	rich	Clu	A.s.n	Thr	. 11.	' دا	Val	GC(Ve	1 10		GGĆ GIY	ACC Thr	CCT	CAA	108	
391 1081	ATC	GAA	AC	A A1	۸ ,	۸۸C	GGA	GGA A	CG	GGA	٨σ٦	GG.	۸ G	AC 4	4C('	CAT				•		Gly	GIn	106	
1141							-	,		٠.,	JU	City	A.	4.	ling.	His	. Pru	A				ATC He	TCT Ser	314 38	-
		CTT Lev	GA.	A GC Gly	ir i	ATA /	14A .ys	GAA A Giu a	GA	AAC	ATG	AA(רו :	π. ι	λC.		GA	۸ (۲	rc c	cr	TCC	۸٤٦	TAT	120	ю
							•		•	vau	MCI	ı.yı	Ph	c /	.sp	Glu	Clu						Туг	40	

Figure:.5a

1201 GAG GAG TAC ATA AAA AAG ATG AGA GAA ACA GAG GAA TAT AAA CCC AGA ACC CAC Tr T TGG Glu Glu Tyr Itc Lys Lys Met Arg Glu Thr Glu Glu Tyr Lys 1200 Pres AIE l'hr Τm 4.20 1261' GGA ACG GTC ATA AAA CCG AAA CTC CCA GAG AAT TTC כדכ TCA GAA AAA GAG AAG 421 Gly Thr 444 Val lic Lys Pro 132D Lys Leu Pro Cin Scr Ciu Glu Lys Lys 440 1321 CCT CCA AAG AAA AAC GAT כוד פנא פוד כוד כופ ATC AGT AGG ATC TCC CCT GAG GGA 441 TAC Pro Pro Lys Lys Asp Val Ala 1380 Val Val lic -Ser Arg He Ciy Clu Cly Tyr 460 GAC AGA AAG CCG GTG AAA GGT GAC TTC TAC CTC TCC GAT GAC GAG CTG GAA Asp Arg Lys Pro Val CTC ATA Lys 1440 Gly Asp Pho Tyr Leu Ser Asp Asp _ Glu Leu He Lys 480 1441 ACC GTC TCG AAA GAA TTC CAC GAT CAG GGT AAG AAA GTT GTG CTT CTG Thr Val Ser Lys Glu Phe AAC ATC GGA His Asp Gin Gly Lys 1500 Lys Val Val Leu lte Gly 500 1301 AGT CCC ATC GAA GTC GCA AGC TGG AGA GAC CTT GTG GAT GGA ATT CTT CTC CTC Ser Pro Ile Glu Val Ala Ser Trp Arg Asp Leu TGG CAG 1560 Val Gly He . Leu Val Trp Gin 520 1561 GCG GGA CAG GAG ATG GGA AGA ATA GTG GCC GAT GTT CTT GTG GGA AAG Ala Gly ATT ccc Gin Glu Met TCC Gly 1620 Arg Ile Val Air Αsp Vaj Val Gly Asn Pro Ser 540 1621 GGA AAA CTT CCA ACG ACC TTC CCG AAG GAT TAC TCG GAC GIT CCA TCC 541 Gly Lys Leu Pro TGG ACG TTC CCA Thr Thr 1680 Phe Pro Lys Αф A.Sp Val Pro \$ = 7 Ттр The Pnc Pro 560 1681 GGA GAG CCA AAG GAC AAT CCG CAA AGA GTG GTG TAC GAG GAA GAC ATC 561, Gly Gle Pro Lys TAC GTG GGA TAC 1740 Azn Pro Gin Arg Val Val Tyr Clu Clu Αsp Тут Cily . Tyr 580 1741 AGG TAC TAC GAC ACC TTC GGT GTG GAA CCT GCC TAC GAA TTC GGC TAC GGC CTC Arg Tyr TCT TAC Tyr Asp Thr 1800 Phe Gly Val Giu Pro Тут Glu Pnc Gly Tyr Gly Scr Tyr 600 ACA ANG TIT GAN TAC ANN GAT TIN ANN ATC GCT ATC GAC GGT GAG ACG CTC The Lys Phe Glu AGA Tyr CTG TCG 1860 Lys Asp Leu Lys Ħε lte Asp Gly Thr Gly Lev Arz Val Ser 620 THE ACG ATC ACA AAC ACT GGG GAC AGA GCT GGA AAG GAA GTC TCA CAG 621 Tyr Thr. He Thr Asn Thr CTC TAC ATC 1920 Gly Asp Arg Ala Gly Lys Glu Val Ser Val Gin Tyr lle Lys 640 1921 GCT CCA AMA GGA AMA ATA GAC AMA CCC TTC CAG GAG CTG Pro Lys Gly Lys lle GCG 777 CAC ACA 1980 ** ** Asp Lys Pro Phe Gln Lev Lvs His Lys Thr Lys 660 CTT TTG AAC CCG GGT GAA TCA GAA GAA ATC TCC TTG GAA ATT 661 CCT CTC AGA GAT GCG CTT 2040 Asn Pro Cly Giu Clu Clu He Ser Clu Pro Lev Arg. ASD Lev Ala 680 2041 AGT TTC GAT GGG AAA GAA TGG GTT GTC GAG TCA GGA GAA TAC GAG 180 Clu AGG GTC CCT GCA 2100 Asp Gly Lys Trp Vat Val Clu Ser Gly Glu Tyt Giv Arg Val 700 Cly Ala 2101 TCT TCG AGG GAT ATA AGG TTG AGA GAT ATT TIT CTG GTT GAG GGA GAG AAG AGA TTC 2160 Arg Asp Arg Leu Arg Asp He Val Glu Gly Glu Lys 720 ATE Lys 2161 CCA TGA 2166 Pro End 722

Figure 56(Continued)

THERMOCOCCUS AEDII12RA GLYCOSIDASE ATC ATC CAC TOC CCG GTT AAA CCG ATT ATA TCT GAG GCT CGC GGC ATA AUC ATC ACA ATA Het lie His Cys Pro Val Lys Gly He lie Ser Glu Ala Arg Gly He Thr He Thr He 61 CAT TTA AGT TIT CAA GOC CAA ATA AAT TTG GTG AAT GCT ATG ATT GTC TIT CCG GAG 20 Asp Leu Ser Phe Gin Gly Gin Ile Asn Asn Leu Val Asn Ale Met Ile Val Phe Pro Glu 120 121 TTC TTC CTC TTT GGA ACC GCC ACA TCT TCT CAT CAG ATC GAG GGA GAT AAT AAA TGG AAC 40 41 Phe Phe Leu Phe Gly Thr Ala Thr Ser Ser His Gln Ile Glu Gly Asp Asn Lys Trp Asn 180 CAC TOG TOG TAT TAT GAG GAG ATA COT AAG CTC CCC TAC AAA TCC GGT AAA GCC TGC AAT 60 Asp Trp Trp Tyr Glu Glu Ile Gly Lys Leu Pro Tyr Lys Ser Gly Lys Ale Cys Asn 240 CAC TOG GAG CTT TAC AGG GAA GAT ATA GAG CTA ATG GCA CAG CTC GGC TAC AAT GCC TAC 241 RO. His Trp Glu Leu Tyr Arg Glu Asp Ile Glu Leu Het Ala Gln Leu Gly Tyr Asn Ala Tyr 300 301 CGC TIT TCG ATA GAG TGG AGC CGT CTC TTC CCG GAA GAG GGC AAA TTC AAT GAA GAA GCC 100 Arg Phe Ser Ile Glu Trp Ser Arg Leu Phe Pro Glu Glu Gly Lys Phe Asn Glu Glu Ala 360 361 TTC AAC CGC TAC CGT GAA ATA ATT GAA ATC CTC CTT GAG AAG GGG ATT ACT CCA AAC GTT 120 Phe Asn Arg Tyr Arg Glu Ile Ile Glu Ile Leu Leu Glu Lys Gly Ile Thr Pro Asn Val 420 ACA CTG CAC CAC TTC ACA TCA CCG CTG TGG TTC ATG CGG AAG GCA GGC TTT TTG AAG GAA 141 Thr Leu His His Phe Thr Ser Pro Leu Trp Phe Het Arg Lys Gly Gly Phe Leu Lys Glu 480 GAA AAC CTC AAG TAC TGG GAG CAG TAC GTT GAT AAA GCC GCG GAG CTC CTC AAG GGA GTC 160 Clu Asn Leu Lys Tyr Trp Glu Gin Tyr Val Asp Lys Ala Ala Glu Leu Leu Lys Gly Val 161 . 540 541 ANG CTT GTA GCT ACA TTC AND GAG CCG ATG GTC TAT GTT ATG ATG GGC TAC CTC ACA GCC 180 181 Lys Leu Val Ala Thr Phe Asn Glu Pro Het Val Tyr Val Het Het Gly Tyr Leu Thr Ala 601 THE TOO CEE CEE TTE ATE AND AGT CEE TIT ANA GEE TIT AND GIT GEE GEN AND CTC CIT Tyr Trp Pro Pro Phe Ile Lys Ser Pro Phe Lys Ala Phe Lys Val Ala Ala Asn Leu Leu 201 660 ANG SCC CAT SCA ATG SCA TAT GAT ATC CTC CAT GGT ANC TIT GAT GTG GGG ATA GTT ANA 661 220 Lys Ala His Ala Het Ala Tyr Asp Ile Leu His Gly Asn Phe Asp Val Gly Ile Val Lys 221 720 AND ATO COO ATA ATO CTO COT GOA AGO AND AGA GAG ANA GAG GTA GAA GOT GOO CAA AAG 240 Asn Ile Pro Ile Het Leu Pro Ala Ser Asn Arg Clu Lys Asp Val Glu Ala Ala Gln Lys 780 GCG GAT AND CTC TIT AND TGG AND TTC CTT GAT GCA ATA TGG AGO GGA ANA TAT ANA GGA 781 Ala Asp Asn Leu Phe Asn Trp Asn Phe Leu Asp Ala Ile Trp Ser Gly Lys Tyr Lys Gly 261 B40 GCT TIT GGA ACT TAC ANA ACT COA GAA AGC CAT GCA GAC TIC ATA GGG ATA AAC TAC TAC 841 280 Ala Phe Gly Thr Tyr Lys Thr Pro Glu Ser Asp Ala Asp Phe Ile Gly Ile Asn Tyr Tyr ACA GCC AGC GAG GTA AGG CAT AGC TGG AAT CCG CTA AAG TIT TTC TTC GAT GCC AAG CTT 300 Thr Ala Ser Glu Val Arg His Ser Trp Asn Pro Leu Lys Phe Phe Phe Asp Ala Lys Leu 960 GCA GAC TTA AGC GAG AGA AAA ACA GAT ATG GGT TGG AGT GTC TAT CCA AAG GGC ATA TAC 961 320 Ala Asp Leu Ser Glu Arg Lys Thr Asp Het Gly Trp Ser Val Tyr Pro Lys Gly Ile Tyr 321. 1020 1021 GAA GCT ATA GCA AAG GTT TCA CAC TAC GGA AAG CCA ATG TAC ATC ACG GAA AAC GGG ATA 340 Glu Ala Ile Ala Lys Val Ser His Tyr Gly Lys Pro Het Tyr Ile Thr Glu Asn Gly Ile 1080 GCT ACC TTA GAC GAT GAG TGG AGG ATA GAG TTT ATC ATC CAG CAC CTC CAG TAC GTT CAC 361 Ala Thr Leu Asp Amp Glu Trp Arg Ile Glu Phe Ile Ile Gln His Leu Gln Tyr Val His 1140 1141 ANA GCC TTA ANC GAT GGC TIT GAC TTG AGA GGC TAC TTC TAT TGG TCT TIT ATG GAT ANC 380 Lys Ala Leu Asn Asp Gly Phe Asp Leu Arg Gly Tyr Phe Tyr Trp Ser Phe Het Asp Asn 1200 1201 THE GAG TEG CET GAG CET TIT ACA CEA CEC TIT GEG CTG GAC GAC GAC TAC ACG ACC 400 401 Phe Glu Trp Ala Glu Gly Phe Arg Pro Arg Phe Gly Leu Val Glu Val Asp Tyr Thr Thr TTC ANG AGG AGA CCG AGA ANG AGT GCT TAC ATA TAT GGA GAA ATT GCA AGG GAA ANG ANA 1261 420 Phe Lys Arg Arg Pro Arg Lys Ser Ala Tyr 11e Tyr Gly Glu 11e Ala Arg Glu Lys Lys 1320 1321 ATA AAA GAC GAA CTG CTG GCA AAG TAT GGG CTT CCG GAG CTA TGA 440 441 Ile Lys Asp Glu Leu Leu Ala Lys Tyr Gly Leu Pro Glu Leu End 455

Figure 6

THERMOCOCCUS CHITONOPHAGUS GLYCOSIDASE - 22G COMPLETE SEQUENCE - 9/95

I TIC CTT SCA COO	
1 TTG CTT CCA GAG AAC TTT CTC TGG GGA GTT TCA CAG TCC GGA TTC CAG TTT GAA 1 Het Leu Pro Glu Asn Phe Leu Trp Gly Val Ser Gln Ser Gly Phe Glo Da	
1 Het Leu Pro Glu Asn Phe Leu Trp Gly Val Ser Gln Ser Gly Phe Gln Phe Glu 61 GAC AGA CTC ACC ACC TO THE GLOVE TREE CONTROL TO THE GOVERNMENT TO THE GOVERNME	ATC GUC 60
61 GIC NCA CTC AND	Met Cly 20
ASP AND LEU AND AND HIS ILE ASP PRO ASE THE ASP THE TAC TOO CTA ACA I	CAT CAA 120
21 ASP Arg Leu Arg Arg His Ile ASP Pro Asn Thr Asp Trp Trp Tyr Trp Val Arg 121 TAT AAT ATC AAA AND CO	Asp Clu 40
121 TAT AAT ATC AAA AAA GGA CTA GTA AGT GGG GAT CTT CCC GAA GAC GGT ATA AAT T 41 Tyr Asn Ile Lys Lys Gly Leu Val Ser Gly Asp Leu Pro Gly Asp Cly 7	
Tyr Ash fle Lys Lys Gly Leu Val Ser Gly Ash Lan Der GAA GAC GCT ATA AAT T	TCA TAT 180
41 Tyr Asn Ile Lys Lys Gly Leu Val Ser Gly Asp Leu Pro Glu Asp Gly Ile Asn S	er Tyr 60
181 GAA TTA TAT GAG AGA GAC CAA GAA ATT GCA AAG GAT TTA GGG CTC AAC ACA TAT A 61 Glu Leu Tyr Glu Arg Asp Gln Glu Ile Ala Lys Asp Leu Gly Leu Arg Ara TAT A	,-
61 Glu Leu Tyr Glu Arg Asp Gln Glu Ile Ala Lys Asp Leu Gly Leu Asn Thr Tyr A	GG ATC 240
The Bys Asp Leu Gly Leu Asn Thr Tyr A	rg Ile 80
241 GGA ATT GAA TGG AGC AGA GTA TTT CCA TGG CCA ACG ACT TTT GTC GAC GTG GAG T 81 Gly Ile Glu Trp Ser Arg Val Phe Pro Trp Pro Thr Thr Phe Val Arg Val	-9 116 80
81 Gly Ile Glu Trp Ser Arg Val Phe Pro Trp Pro Thr Thr Phe Val Asp Val Glu Tr	AT CAA
The Fit iff pro Thr Thr Phe Val Asp Val Glu T	AT GAA 300
101 ATT GAT GAG TCT TAC GGG TTG GTA AAG GAT GTG AAG ATT TCT AAA GAC GCA TTA GJ 101 Ile Asp Glu Ser Tyr Gly Leu Val Lys Asp Val Lys Ile Ser Lys Asp Val Cys Asp Val	yr Glu 100
101 The Asp Glu Ser Tyr Gly Leu Val Lys Asp Val Lys The Ser Lys Asp Ala Leu Gl 361 CTT GAT GAT ATC COT LLC	
121 Leu Asp Glu Ile Ala Asn Gln Arg Glu Ile Ile Tyr Tyr Arg Asn Leu Ile Asn Se	
421 AGA AAG AGG GGT TIT AAG GTA ATA CTA AAC CTA AAT CAT TIT ACC CTC CCA ATA TG 141 Arg Lys Arg Gly Phe Lys Val Ile Leu Asn Leu Asn His Phe Thr Leu Pro Ile Tr	
ory Fre Lys Val Ile Leu Asn Leu Asn His Phe Thr Leu Bre Th	CLI 480
481 CAT GAT CCT ATC GAA TCT AGA GAA AAA GCC CTG ACC AAT AAG AGA AAC GGA TGG GT. 161 His Asp Pro Ile Glu Ser Arg Glu Lys Ala Leu Thr Asn Lys Arg Asn Gly Trp Va.	
The Glu Ser Arg Glu Lys Ala Leu Thr Asn Lys her the Glu	A AGC 540
541 GAA AGG AGG AGG GEO CONTRACTOR OF THE CASE OF THE VALUE OF THE VAL	l Ser 180
541 GAA AGG AGT GTT ATA GAG TIT GCA ANA TIT GCC GCG TAT TTA GCA TAT ANA TTC GGJ 181 Glu Arg Ser Val Ile Glu Phe Ala Lys Phe Ala Ala Tor Lou Ala CAT TAT ANA TTC GGJ	
181 Glu Arg Ser Val Ile Glu Phe Ala Lys Phe Ala Ala Tyr Leu Ala Tyr Lys Phe Gly	€ CYC 800
601 ATA CTA CAS AND	/ A sp 200
201 Ile Val Asp Het Trp Ser Thr Phe Ash Glu Pro Het Val Val Val Com GGG TAT	
201 The Val Asp Met Trp Ser Thr Phe Ash Glu Pro Met Val Val Ala Glu Leu Gly Tyr	TTA 660
661 CCC CCC TO THE GIV TY	Leu 220
661 GCC CCA TAC TCA GCA TTC CCC CCG GGA GTC ATG AAT CCA GAA GCA GCA AAG TTA GTT 221 Ala Pro Tyr Ser Gly Phe Pro Pro Gly Val Het Asp Pro Gly Ala CCA AAG TTA GTT	
Ala Pro Tyr Ser Cly Phe Pro Pro Cly Val Het Asp Pro Cly Ser AAG TTA GTT	ATG 720
TO GIU AIA AIA TUE TAN TIET	
CAN CAT ATT: 110 COO OL	
721 CTA CAT ATG ATA AAC GCC CAT GCT TTA GCA TAT AGG ATG ATA AAG AAA TTT GAC AGA 241 Leu His Het Ile Asn Ala His Ala Leu Ala Tyr Arg Het Ile Lys Lys Phe Asp Arg 781 All COM DIA	XXX 780
THE LYS LVE Pha 1cm 1-a	1
TO AND ULT GAT CC) Cas mas asset	
781 ANA GCT GAT CCA GAA TCA ANA GAA CCA GCT GAA ATA GGA ATT ATA TAC ANT AAC ATC 261 Lys Ala Asp Pro Glu Ser Lys Glu Pro Ala Glu Ile Gly Ile Ile Tyr Asn Asn Ile 841 GTC ACC DUR GROUP COMMON CO	GGC. 840
VAI ALE TVY len las TI	
841 GTC ACA TAT CCG TTT AAT CCG AAA GAC TCA AAG GAT CTA CAA GCA TCC GAT AAT GCC 281 Val Thr Tyr Pro Phe Asn Pro Lys Asp Ser Lys Asp Leu Gln Ala Ser Asp Asn Ala	AAT 900
THE DEE ASD	100
TTT TO THE EAC ACT CON CON	
301 Phe Phe His Ser Gly Lou Phe Lou Thr Ala Ile His Arg Gly Lys Lou Asn Ile Glu	TTT 960
The state of the s	* \
321 ASP Gly Glu Thr Phe Val Tyr Lou PAT TTA AAG GGC AAT GAT TGG CTG GGA GTG	AAT 1020
The state of the s	
341 Tyr Tyr Thr Arg Glu Val Val Lys Tyr Gln Asp Pro Het Phe Pro Ser Ile Pro Leu 1	ATA 1080
AUC TIC AAG GGC COM OCC ASS	
361 SET Php ING Cly Val	
TATA GOT AAT CITY COME AND	
1141 GGT AAT CCT GTT AGT GAC ATT GGA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT A 381 Gly Asn Pro Val Ser Asp Ile Gly Trp Glu Val TVP Pro Law Gly Age TGT GAC TCT A	
381 Gly Asn Pro Val Ser Asp Ile Gly Trp Glu Val Tyr Pro Lys Gly Het Tyr Asp Ser I	TA 1200
1201 CTA CCT CCC 112	le 400
1201 GTA GCT GCC AAT GAA TAT GGA GTT CCT GTA TAC GTA ACA GAA AAC GGA ATA GCA GAT T 401 Val Ala Ala Asn Glu Tyr Gly Val Pro Val Tyr Val The Gly Aca GA GA GA GAT T	
401 Val Ala Ala Asn Glu Tyr Gly Val Pro Val Tyr Val Thr Glu Asn Gly Ile Ala Asp 5	CA 1260
The The DAY ASD GIV TIGHTS AND A	430
ANN GAT GTA TTA ACC COC TATA	
1261 AAA GAT GTA TTA AGG CCC TAT TAC ATC GCA TCT CAC ATT GAA GCC ATG GAA GAG GCT TAC Lys Asp Val Leu Arg Pro Tyr Tyr Ile Ala Ser His Ile Glu Ala Het Glu Glu Ala TC	AC 1320
The set his ite Giu Ala Her Chi chi ala me	440

Figure 7a

1121	CX	AAT	. CC1	. TA7	. CAC	. ~~														TOG	
441	Clu	Asn	GIV	7			AGA	CCY	TAC	TTA	CAC	TCC	GCA	₩.							
			,	. 72	Asp	, API	Arg	Gly	Tyr	Leu	Hi.	Ten	11-	: '^	ACC	GAT	AAT	TAC	GVV	TCC	1 180
1381	CCC	TTA	CCC	~~~										240	inr	VED	Asn	TYT	Clu	Trp	460
461	Ala	Len	Clar	2,	707	ATC	ACC	111	CCC	TIC	TAC	GAA	~								
461 1441			Oly	rne	λIg	Met	Arg	Phe	Glv	1.00	7	Class		VVC	110	ATA	ACC	AAA	CAC	AGA	1440
																					-
1441	w	CCC	AGG	AAA	AAG		~		_									-,-		vid	480
481	Lys	Pro	Arm	Lva	****	701	CIA	AGA	CTA	TTC	AGA	GYC	ATA	CT-7						ACA	
	•		••••	2 73	LY3	Ser	Val.	Arg	Val	Phe	Ara	Glu	Tla		21.	AAT.	AAT	CCC	CTA	ACA	1500
						GAG								VAI	He	Asn	Asn	Gly	Leu	Thr	500
501	Ser	Asn	Tia		~~~	CAC	ATC	TTA	CAC	GAC	CCC	TAG	15	16							
			114	vià	Lys	Clu	Ile	Leu	Clu	Clu	Glv	End	< 1						•		

Figure 7b(Continued)

PYROCOCCUS FURIOSUS GLYCOSIDASE - 7G1 COMULETE GENE SEQUENCE - 10/95

				_									. 32	COL	MCX	- 1	0/95	i								
	1	ATG	TT	: ć:	TC	AA /	4 .G	TTC	CIT	TG	6 60	-			٥.,				TT C							
	•	net	Phe	• P1	C	lu 1	Lys	Phe	Leu	Tr	D G1	vì	/a1	Ala	61	TC	C GC	ST T	II (CAC	TTT	GA	A A	TG.	GGG	60
	61 (CAT	AA			~~ ·													.,.		Pne	CI	א ט	e t	Civ	20
	21 ; 21 ;	qe۸	Lys	Le	- A	- C	IGG	AAT	ATT	CX	: AC	T A	AC	ACT	GAT	TG	ם דר	-	AC I						•	
		•	-,-			19 7	ug.	Asn	Ila	As;	o Th	ΓÀ	Gn.	Thr	Asp	Tr	D Tr	א פי	AC I		GTA	AG	G G	AT.	AAG.	120
1	21 /	ACA	AAT		3 6										-		•			r p	AGI	AI:	3 A:	30	Lys	40
·	41 ;	lte	מפא	11	÷ G	lu L	YS	Gl v	Lau	Val	· AG	T G	GA (CAT	CII	, cc	ב כא	نت عا	AG G	GG :	1 TT					
11	91 6						•	,		. 41		r C	T Y	Asp	Leu	Pr	o - G1	ង G	lu G	ly	lle	An	. ~		rac	180
	51 0	:1,,	t TI	TA	I C	AC A	אה :	SAC	CAT	GAG	λT	rc	CA:	CZ	À			_	lu G TT A	_					y	60
	•		200	- y	r G	ru L	λ2)	Asp	His	Glu	Ile	e A	2 a 2	Lra	Lvs	7.0	. 66	T C	IT A	AT (CT	TAC	: Ac	λ,	ATA	240
24	11 0	CC	ATA	C2/											•			,		J., /	~ a	TY	λı	a 3	lle	80
ε	11 6	ìγ	Il.	Gli	u Tz	70 S	er 1	\ T.G	ATA	TTC	SC	· I	GG (CA	ACG	λCA	TT	T A:	T G	AT 6				_		
						-	/	u y	TTE	rne	Pro	> T:	rp E	, LO	The	Thi	Ph	e Il	e A	74 L	122	GAT	TA	7 7	'ec	300
30	I T	XI	TKA	€\$1	V IC	A T	A TA	VAC	CIT	ATE	(A)		·								41	AG p	· Ty	r S	ez	100
10	ı T	yr 2	720	G1:	1 Se	r T	ys A	תבו	Leu	Ile	Glu		13 U	TA	AAG	ATC	AC	C AA	ie A: iG Gi 's A:	K DA	CT	TIG	GA	c c	3.5	3.50
36	1 T		· i =	~									۷ د	47	rys	lle	Th	- Ly	'3 እ:	P I	hr	Leu	GI	uG	lu	360
12	ii	eu J	, *D	CAC	AT	C CC	כ א	AC .	NAG	AGG	GAG	GI	.c =	cc	TAC	ተ ът	N C C		'3 A: A GI							120
		,	io b	310	1 11	e Aj	ra y	מכ.	Lys	Arg	Glu	Va	1 A	la	Tyr	Tyr	AGG	, C	A G1	CA	TA	AAC	AG	c	75	420
42	1 A	33 2		3 50												-		,	- "		TE	ASD	5 e.	: L	eu	140
14	ند د	: Ţ S	e=	Lys	G.	y 25	6 7	~	7-7	ATA	GTT	λA	TC	TA.	L LL	CAC	770	: AC	C CT F Le	7 0	י בי	T 7 T				
				-						1 = 6	Λ4.	3.0	n L	eп.	` ==	X. s	Spe	Th	r Le	u P	~A	LV.	TGC	• T	T5	480
48	. C2		λT	CCC	AT:	I CA	္ မ	CT 3	GG	وكرو) AGG		C +-						F Le G XX	-		. 3 -	1	, T.	ŧu	. 160
16	1 7.2	. 3 A	зp	Pro	Il	e G1	u A	la	rg .	Glu	Ara	Al	G 1.	EA A	ACT .	XXI	λλc	AG	g aa	C 6:	sc :	155	GTI	. ه	.c	540
543			C.A.	.~.			_		•							AJN	Lys	Ar	5 AA 5 A3	n G	ly :	rp	Val	λ	תנ	180
191	2.	0 A		ار الراك مرحد	Val	F AT	Д С.	VG I	TT (GCA	AAG	TA:	T G(c c	SCT '	TAC	ATA	cc	TA		_					
•			•		,	- 11	e 6.	ru P	he ;	Ala	Lys	Ty:	ב א	La J	المك	Tyr	Ile	נגג	TA:	I	\G :	III.	CCY	تی .	T	600
601) AI	A G	TG (SAT	ATG	TG	G Ac		ra .							•			y	,	/3 }	he	Gly	λ	Þ	200
201	. 11	e V	al ;	なら	Met	TI	P Se	:: T	hr i	Phe	AAT	CAC	3 CC	T 2	ITG (STG	CII	GII	TY:	e CI	7 6	CC	ቸኳሮ	C T	· n	c-0
661	5		~ .				_	_			٠	011	ı Pı	- P	ie E	Val	Val	Val	G11	ı Le	u G	lv	Tvr	1 -	<u></u>	660 220
221	Αl	2 P		LAC.	TCT	CCC	- 11	.c c	CI C	CA (GGG	GTI	CT	'A A	AT C	-CA-	C3.C		. CC			-			_	
		- • •		. 7 -	ser	(L_)	y Ph	e P	to E	, to	Gly	Va]	عتا	u A	מב.	250	Clu	B) 2	Y) a	· XX	e c	IC	CCC	AT	A	720
721	CI.	T CZ	\C #	170	200														مبد	гту	3 L	eu.	AL a	11	e	240
241	Le	נול ט	. 5 }	ier	Ile	Ası	I Al	A (.	AT G	CT	TA	éc.	TA.	A I	GG (AG .	ATA	AAG	AAC Lys	ידיד :	т' с	۱.	N ~~	~ >	_	
781										u.a.	Leu .	AL A	ту	x A	13 0	ln	Ile	Lys	Lys	Ph	e A	3D '	The	(2)		780
261	T	GC	TC	AT	AAG	GAT	TC	TA	W G	AG C	CT.	CC N	- 2	> ~					TAC					01	_	260
-01	Ly:	A.	.a A	3p	Lys	Asp	Se.	r Ly	/3 G	lu I	20	AL A	G1	M G	11 6	GI :	ATA	ATT	TAC	AA	CA	AC J	TT	CG	4	840
841	GIT	GC	TT	TA				_								-, .			1 7 2	A.51	n A	מנ	[]e	G1	1	280
251	لدلا	Al	a T	V.F	Pro	TV	GA.	I CO	CA	AC G	AI :	TCC	AA	S GJ	AT G	TI ;	AAG	GCA	GCA Ala							
				-		2,3	~31	2 22	o Y	א מכ	ap :	Ser	Ly	9 A:	p v	al 1	Lys	Ala.	Ala	GI.		VC 6	AC	AAC		900
901 301	TTC	1.	C C	AC '	TCA	GGG	CTC	TT	C 7*	rc c	100	~~~					•		CTT	G 1.		in A	Gр	V2L	•	300
301	Phe	Ph	e . H		Ser	Gly	Lei	ı Ph	c P)	ne G	111 3	112	ATA		(C X	AA C	GA.	እእኢ	CTT Leu	AAT	רא י	A G	ac.	TTT		960
961	GAC	cc	T (***	nı.	. 9 . L	уз С	ily .	Lys	Leu	Asr	ı Il	e G	lu	Phe		320
321	Asp	GI	4 (m/		RCG	III	ATA	GA	I GC	C C	CC 1	AT	CTA	. AA	c c	- n	AT (C 3 C	TCC							
	•		, .		inr	rne	Ile	: λ 5	P A.	a P	ro T	'yr	Leu	Ly	3 G	lvä	OL I	220	TGG Trp	ATA	GC	G C	II.	AAI		1020
021	TAC	TAC	- 20	2. 1										•					LLP	TTE	G 7	y V	aŀ.	חבא		340
341	Tyr	Ty	T	ır J	وعا	Glu	Val	Va	1 AC	C T	RT C	AC	CYA	CC	አ እን	rc T	TT (CCT	TCA Ser	ATC	כר	a c	TC 1			
081	300				•			٠	• •••		yr G	τn	Clu	Pr	o Me	t P	he l	Pro	Ser	Ile	P=	0 L	-11 '	11-		1080 360
361	The	117	A.4	G	GA	GTT	CAA	CC	KT J	T GC	ic t	A.T	CCC	TC	c							_			•	,,,,
	4	- 416	Ly	3 C	ı y	Val	Gln	Gl	Ty	r C	YT	yΣ	ملم	CV	- AG	 	CT C	-GA	Ser ACT Thr	CIG	TC.	ند د	AG (ΞAΙ	1	140
141	GAC	AGA	CC	c -										-		-		,		F 9 ff	5 e	= L	ys y	de/		90
381	GAC Asp	Arg	2.	- U	ובי הו	MUC Sar	WAC	ATA	CC	A TO	:C G	AA	CTC	TA	r cc	A G	AG G	cc	k TC	T > C						
	CT: Yap	•	. •	- •		-61	vab	116	G1	y ::	P G	lu	Leu	Ty	r Pr	o G	lu G	ly	Met	TAC	ωA.		A A	TA		200
203	GT: Val	حمه	C.	T C.	AC 2	W.G	TAC		مطمنا	T ~~				_				~ 1		. Y L	~ -	- >6	:r 1	1 e	4	00
401	Val	Clu	A.	a H	15 I	.ys	Ty	GIV	۷۷.	ים נכ יים נ	A 67	T 1	TAC	GTO	λC	CC	IG A	AC (GGA	ATA	GCC	; ca	T T	CΛ	,	260
						•	• -	,	- 4.	- +1	J V.	• 1	ı yı	Val	Th	r G	Lu A	sn (51 y	lle	Ala	ı As	ر 2 و	er.		20
																			-						•	

Figure 8a

Lys Asp Ile													1320
													1380
													1440 480
ATT CCC AGG					ATA Ile	GTA Val	GCC Ala	TAK ark	AAT	GGT	GTT	ACG	1500
ANA ANG ATT Lys Lys Ile					33					J.y		inr	500

Figure 8b(Continued)

Bankia gouldi endoglucanase (37071)

									•				• • •					
9		•	18			27			~ ~									
5 .	ATY								36			45			54			
-		, 700	v v13	· CG	r Tr	A GC	ם אכו	3 CM	C GC	3 CIC	TGC	: GCX	CCC	CITC	100	CCA		
	net	. Ar	y Ile	Ar;	Le	u Al	a Thi	Le	. A1	1.00	0~	27-			٦٠٠	CCA	GIC	, YCC
	_				•						. cya	. WIR	· VIA	ren	Ser	Pro	Val	ACC
	•		63															
	-					7	2		8.	l ·		90)		90			
			L GAT	, YY3	CI	y yc	CTD T	ו כאו	XX	: GAC	GCC	CAC	-			***	·	TÓR
	Phe	: Ala	ds _k) Asi	Va.	l Th	r Val	(3)	. 71.			-		CGI	AAA	Lys	CLC	ATC
			•							ىرىم.	- ALA	veb	CIA	CJA	Lys	Lys	Leu	Ila
			117													-	-	
	١.٠٠					12	5		135	i		144			163			
	· ALC	CG	CCC	CLI	TA	C GG	: ATC	: AAT	. ***	TCC	220				133	XCC		162
	Ser	Arg	Ala	Leu	TV	c G13	/ Mer	1 or		· c		~~	- CAA	ALC	CLL	ACC Thr	GAT	ACT
•								, ,,,,,,,,	, ABI	. Ser	ASD	YIS	Glu	Ser	Leu	Thr	Asn	The
			171			180)		189			198			207			
	CAC	TGG	CAG	CGI	TI	CGC	CAT	GCA	GGT	CAC:	CCC		~~~		207	λλτ		216
	Asp	Trp	Gln	λισ	Phe	. Arc) den	324	01		-0-	AIG	CIG	CGG	GYY	λλΤ	GGC	CGC
		_					الوكار ا	VIG	CIA	ATT	λīg	Met	Leu	Arg	Glu	λλΤ λ32	Glv	Glv
	•																3	- -y
			225		•	234	ı		243		•	252			261		•	
	AAC	XXC	YCC	YCC	XXX	TAT	AAC	TCC	CAR	-	CIC				70 T	CCC (270
	λsn	λsn	Ser	Thr	Tive	400	>	~	C)-	-	مد	CIG	AGC	AGT	CAT	ccc (GAT	TGG
					-, -		****	Trp	GID	rea	His	Tan	Ser	Ser	His	Pro 2	ARD :	
			279			288			297			306			315			
	TAC	YYC	λλT	GIC	TAC	GCC	CCC	AAC	AAC	110	W-C				313			324
	$\mathbf{I}^{\mathbf{X}\mathbf{I}}$	מבע	λsn	Val	Tyr	Ala	Glv	100		1	200	CALL	AAL	CGG	GTA	CCC (ITG .	ATT
					-3-		GIY	VBII	ABD	ASD	TTP	Vab	Asn .	, עבא	Val .	Ala 1	eu :	Il e
			333															
	~~~					342			351			360			3.60			
	CAG	CYY	YYC	CLC	CCC	CCC	CCC	GAC	ACC	ATY				~~~				378
	Gln	Glu	Asn	Leu	Pro	Glv	Ala	len	The	V		~~	TIC	CAG (	erc .	ATC C	ET 1	<b>NG</b>
							•	رومد	1111	ACE	TED .	ΥΤΦ	Phe (	Glm 1	Leu :	Ile G	lv i	Lvs
			387															
			361			396			405			414			123			
	GIC .	GCG	GCG .	ACT	ICI	GCC	TAC	λλC	TIT	AAC (			C			CAG I	•	132
	Val .	Ala	Ala	Thr	Ser	Ala	Tv-	len	Dh.	}			GAA .	TTC 3	NG (	LAG I	<b>CG</b> (	'AA
	Val .								21112	. تنجه،	asp :	المتا	Glu 1	ye ac.	rev (	Sla S	er G	ln
			441															
	TCC :	~~	111			450			459		4	168		4	77			0.0
	TCG :	100	YCC (	GGC	GIC	CCT	CAG .	AAT (	CTC	GCT (	360 (	355 (	ם מכב		···			00
	IID ;	طعه	Thr (	Gly '	Val -	λla	Gln	Asn 1	1.011	A 7 a . C	2720					MI C	TG G	AC
	Trp:										ora (	зτλ (	TA C	lu P	LO Y	usn L	eu A	3p
			495											•				
	GGC 4					504			513		5	<b>32</b> 2		5	31		E	40
	GCC (	366	GGC (	SAA (	SCG	CIG	CLI (	SAA (	GGA (	SAC C	CC A	AT .	ىك مىكىت	× ~	<del>~</del>		<u>-</u>	40
	CIA (	sly (	Gly (	ilu i	Nla	Leu	Val (	21u 6	21v 1	lon t	) - 1			AC (	IC A	arg ce	AT T	GG
	Gly (								, .	P	-U 7	mi I	eu T	AI L	eu M	et A	T QE	φ.
			549			558									•		_	-
	TCG C	· .	~~~					-	67		5	76		5	85		. E	1
	TCG C Ser P			MC )	CT	GIG (	GGT A	ATT C	TC C	NC C	AC T	GG T	בו עוב	GC ~	 TD -	٠		- <b>4</b>
	ser F	Lo 3	Ma A	BD 1	Chr '	Val (	Gly I	le I	eu A	an H	ie m		·	SC 0	-A A	AC G	<b>35</b> C.	IC.
	Ser P									-y 41		י איני	ne G	TA A	al Y	sn Gl	y L	574 ·
			503			612												
	acc o	· 'Y'	-	~		012		, 6	21		6	30		63	39			• D
	erà a ecc e		.uu C	GT C	GC )	AAA (	SCC Y	AA T	'AC I	GG A	GT A	TG G	AT A	) C		na		
,	Gly V	al A	ITG A	rg G	ly i	Lys J	la L	VE T	Vr 7	E	L			. u	יים כו	CC 00	C Y	TC .
					~			•	4 - L	-y 3	- M	et A	ab y	av C]	Lu P	Lo CJ	y Il	. e
			57															
	rcc ~	ں ۔ جب	- ·			566		6	75		61	B 4 -		50	1		~~	
	LED A		GC X	cc c	YC (	SAC G	AT G	TA G	TG A	AA GJ			~ ~	~ ~	, <u>,</u>		_ /0	2
7	א פבן	al G	ly T	br H	is )	K gal	ED V	al v	a 3 - 1 -				בים בינ	G G1	Y C	n cy	L LI	ζ.
					-		v	V		ys G.	ra G	IN T	hr Pr	o Va	il Gi	lu As	p Ph	•

Figure 9a

# Bankia gouldi endoglucanese (37GP1) (continued)

CTG CAC ACC TAT TTC GAA ACC GCC AAA AAA GCC CGC GCC AAA TTT CCC GGT ATT 720 Leu His Thr Tyr Phe Glu Thr Ala Lys Lys Ala Arg Ala Lys Phe Pro Cly Ile 774 ANA ATC ACC GGT CCC GCT AAT GAG TGG CAG TGG TAT GCC TGG GGC GGT Lys Ile Thr Gly Pro Val Pro Ala Asn Glu Trp Gln Trp Tyr Ala Trp Gly Gly 819 TTC TCG GTA CCC CAG GAA CAA GGG TTT ATG AGC TGG ATG GAG TAT TTC ATC AAG Phe Ser Val Pro Gln Glu Gln Gly Phe Met Ser Trp Met Glu Tyr Phe Ile Lyr 882 891 CGG GTG TCT GAA GAG CAA CGC GCA AGT GGT GTT CGC CTC CTC GAT GTA CTC GAT Arg Val Ser Glu Glu Gln Arg Ala Ser Gly Val Arg Leu Leu Asp Val Leu Asp 936 CTG CAC TAC TAC CCC GGC GCT TAC AAT GCG GAA GAT ATC GTG CAA TTA CAT CGC Leu His Tyr Tyr Pro Gly Ala Tyr Asn Ala Glu Asp Ile Val Gln Leu His Arg 990 ACG TTC TTC GAC CGC GAC TTT GTT TCA CTG GAT GCC AAC GGG GTG AAA ATG GTA Thr Phe Phe Asp Arg Asp Phe Val Ser Leu Asp Ala Asn Gly Val Low Met Val GAA GGT GGC TGG GAT GAC AGC ATC AAC AAG GAA TAT ATT TTC GGG CGA GTG AAC Glu Gly Gly Trp Asp Asp Ser Ile Asn Lys Glu Tyr Ile Phe Gly Arg Val Asn 1098 CAT TOG CTC GAG GAA TAT ATG GGG CCA GAC CAT GGT GTA ACC CTG GGC TTA ACC Asp Trp Leu Glu Glu Tyr Met Gly Pro Asp His Gly Val Thr Leu Gly Leu Thr 1143 1152 GAA ATG TGC GTG CGC AAT GTG AAT CCG ATG ACT ACC GCC ATC TGG TAT GCC TCC Glu Met Cys Val Arg Asn Val Asn Pro Met Thr Thr Ala Ile Trp Tyr Ala Ser 1197 1206 ATG CTC GGC ACC TTC GCG GAT AAC GGC GTC GAA ATA TTC ACC CCA TGG TGC TGG Met Leu Gly Thr Phe Ala Asp Asn Gly Val Glu Ile Phe Thr Pro Trp Cys Trp 1260 AAC ACC GGA ATG TGG GAA ACA CTC CAC CTC TTC AGC CGC TAC AAC AAA CCT TAT Asn Thr Gly Met Trp Glu Thr Leu His Leu Phe Ser Arg Tyr Asn Lys Pro Tyr 1314 CGG GTC GCC TCC AGC TCC AGT CTT GAA GAG TTT GTC AGC GCC TAC AGC TCC ATT Arg Val Ala Ser Ser Ser Ser Leu Glu Glu Phe Val Ser Ala Tyr Ser Ser Ile 1368 AMC GAA GCA GAA GAC GCC ATG ACG GTA CTT CTG GTG AAT CGT TCC ACT AGC GAG 1377

Asn Glu Ala Glu Asp Ala Met Thr Val Leu Leu Val Asn Arg Ser Thr Ser Glu Figure 9b(Continued)

# Bankia gouldi endoglucanase (37GP1) (continued)

1413 1422 1431 1440 1449 1458
ACC CAC ACC GCC ACT GTC GCT ATC GAC GAT TTC CCA CTG GAT GGC CCC TAC CGC
Thr His Thr Ala Thr Val Ala Ile Asp Asp Phe Pro Leu Asp Gly Pro Tyr Arg

1467 1476 1485 1494 1503 1512
ACC CTG CGC TTA CAC AAC CTG CCG GGG GAG GAA ACC TTC GTA TCT CAC CGA GAC
Thr Leu Arg Leu His Asn Leu Pro Gly Glu Glu Thr Phe Val Ser His Arg Asp

1521 1530 1539 1548 1557 1566
AAC GCC CTG GAA AAA GGT ACA GTG CGC GCC AGC GAC AAT ACG GTA ACA CTG GAG
AEN Ala Leu Glu Lys Gly Thr Val Arg Ala Ser AEP AEN Thr Val Thr Leu Glu

1575 1584 1593 1602 1611
TTG CCC CCT CTG TCC GTT ACT GCA ATA TTG CTC AAG GCC CGG CCC TAA 3*
Leu Pro Pro Leu Ser Val Thr Ala Ile Leu Leu Lys Ala Arg Pro ***

Figure 94 (Continued)

# Thermologa maritima Alpha-galactusidade Complete Gene Sequence (1 + 4 3)

,
5° GTG ATC TOT CO. 18 27 16
5' CTG ATC TGT GGA ATA TIC GGA ANG ACC TTC AGA GAG GGA AGA TTC GTT CT
Val Ile Cys Val Glu Ile Phe Gly Lys Thr Phe Arg Glu Gly Arg Phe Val Le
63 63 63
Lys Glu Lys Asp Pho The Act of the God of Grand and Ath Cac of Grand Cac of Grand and Ath Cac of Grand Cac of
Lys Glu Ive
Lys Glu Lys Asn Phe Thr Val Glu Phe Ala Val Glu Lys Ile His Leu Gly Trp
117 126 135
Lys Ile Ser Cly No. 2126 135 144 153 162  Lys Ile Ser Cly No. 2126 125 125 144 153 162
Lys Ile Ser Gly Arg Val Lys Chy See By
Lys Ile Ser Gly Arg Vel Lys Gly Ser Pro Gly Arg Leu Glu Val Leu Arg Thr
ANA GCA CCG GNA NAG GTA CTT GTG ANC NAC TCG CNG TCC TCG GCA CCG TCC AGG
Lys Ala Pro Glu Lys Val Leu Val Asn Asn Trp Gln Ser Trp Gly Pro Cys Arg
CTG GTC GAT GCC TIT TCT TTC AAA CCA CCT GAA ATA GAT CCG AAC TGG AGA TAC
Val Val Asp Ala Phe Ser Phe Lin Day
Val Val Asp Ala Phe Ser Phe Lys Pro Pro Glu Ile Asp Pro Asm Trp Ary Tyr
ACC GCT TCG GTG GTG CCC GAT GTA CTT GAA AGG AAC CTC CAG AGC GAC TAT TTC
Thr Ala Ser Val Val Pro Asp Val Leu Glu Ary Asm Leu Glm Ser Asp Tyr Phe
GTG GCT GAA GAA GGA AAA GTG TAC GGT TIT CTG AGT TCG AAA ATC GCA CAT CCT
Val Ala Glu Glu Gly Lys Val Tyr Gly Phe Leu Ser Ser Lys Ile Ala Ris Pro
THE THE GET GIG GAA GAT GIG CAN GIT GIG GAA GAT GAT GAA GAA
THE THE GET GIG GAA GAT GGG GAA CIT GTG GCA TAC CITC GAA TAT THE GAT GTC
Phe Phe Ala Val Glu Asp Gly Glu Leu Val Ala Tyr Leu Glu Tyr Phe Asp Val
NN1 160
GAG TTC GAC GAC TTT GTT CCT CTT GAA CCT CTC GTT GTA CTC GAG GAT CCC AAC
Glu Phe Amp Asp Phe Val Pro Leu Glu Pro Leu Val Val Leu Glu Asp Pro Asm
495 504 504 Fro Leu Val Val Leu Glu Asp Pro Asn
ACA CCC CIT CTT CTG GAG AAA TAC CCC CIT CTT CTG GAG AAA TAC CCC CTT CTT CTG CTG CTT CTT CTG GAG AAA TAC CCC CTT CTT CTG CTG CTG CTG CTG CTG CT
THE GET GIT GET AND AND AND AND THE
The Pro Leu Leu Clu Lys Tyr Ala Clu Leu Val Cly Met Glu Asn Ala
399 .866
AGA GTT CCA ANA CAC ACA CCC ACT CCA TCC TCC ACC TCC TAC CAT TAC TTC CTT
Arg Val Pro Lys His The Pro The City The Comment
Arg Val Pro Lys His The Pro The Gly Trp Cyr Ser Trp Tyr His Tyr Phe Leu
·

Figure 10a

# Thermotoga maritima Alpha-galactosidane Complete Gene Sequence $(2,o(\beta))$

														•						
	_		60	3		6	12			62	1			636			6.3	10		
	AT .		: AC	ינ זכ	æ €	VV C	AG A	ונינ'	CK	: W	c v	NÇ (	CIC	λΛ	ב כא	c 00	ຮຸ້ນ	ຮຸ້ນ	AT TT	641 200
λı	D .	ieu	Th	r Tr	p G	lu G	 lu 7	Dir	Les	ı Lv	5 As	·		7.00						e Pro
												••••	<b>∽e</b> u	Ly	, re	1 1/1	a Ly	~ <b>Λ</b> ι	את תו	e Pro
T	C (	SAG	65 65	, : 11	~ C	6 'A G A'	66 Tag	:	CAC	67	5 ~ ~~		٠	684			69	3		702 G CIC
==	-													AAL	CAC	AT	A GG	1 CY	'⊂ 10	c crc
Ph	LER C	ilu	Va.	l Ph	e Gi	in I	le A	æБ	λsp	, YT	a Ty	<b>T</b> (	lu	Lys	As	110	e G1	 У Ле	P In	 P Leu
			71:	i		7:	20			72				720				_	•	
<b>7</b> 1	G )	CA	λG	GG	y Cy	CT	T C	CA	TCG	GI	CY	λG	λC	λTG	GCA	. AAJ	(or	, I AT	λοσ	756 ርኢአ
																				Glu
					•					•		<b>u</b>	14	MEL	مىم	Lyz	s va	rit	e Ala	Glu
λλ	cG	GT	765		~ ~	77 G G	'4 '~ 1'	TTA	<b>~~</b> ~	783		~ ~		792			80:	Ĺ		810
																				1CC
λει	a G	lУ	Phe	· Ile	Pr	o Gl	y I	Le	Trp.	Thr	. VJ	a P	ro	Pbe	Ser	Val	Sex	Gli	a The	Ser
			819			82	8			837				845			055	•	. •	
GA:	C	IX.	TIC	XXC	. CY	y CY	TC	3	CAC	TCG	ப	4 G	TG .	λλG	GRA	AAC	(CZ)	. כאנ	: 000	864 22G
													_							Lys
														uy s	GLU	YZII	GLY	GIL	PTO	Lys
ATC	; cx		873 TAC	AGX	AAr	88: TT	2 7	د . ت		891	3 773	מיני		900	~~~		909		XXX	918
net	ىد :	a '	IYI	yrg	Yar	IXI	א ק	n I	'Az	Lys	Ile	T	~ )	<u>Ua</u>	Leu	yzb	Leu	Ser	Lys	λsp
			927			936	5			945			5	54			963		•	972
GAG	્લ	T (	CIG.	AAC	TCC	cm	TI	<b>C</b> (	AT	CIC	TTC	TC	A 2	CI	CIG	YCY	AAG	λTG	ccc	TAC
Glu	Va	1	Leu	Asn	Trp	Leu	. Pb	 e >	محا	Leu	Phe	Se	 - S	er	Leu	λru	Lva	Met	Gly	TVT
			_		-							-					_		0.1	131
AGG	TΑ		981 [TC	AAG	ATC	990 OAD	TT	r c	т .	999	CC	œ	10 T C	800	طعلت	~,1	1017	~ .	AGA.	026
λtg	13	T E	ne	Lys	Ile	ysb	Ph	= L	en l	Phe	Ala	G1	у	la '	Val .	Pro	СĵЪ	Glu	yra	Lys
			35			1044			10	253			10	62		1	071		1	080
AAG	AA	C 7	TA.	YCY	CCA	ATT	CXC	; C	cc 7	TC	AGA	AA,	A C	CC 1	ATT (	CAC	ACG	ATC	<b>YCY</b>	XXX
Lye	بحخ	n I	le	Thr	Pro	Ile	Glr		la I	he	λrg	Lys	 3 G	: ly :	ile (	5lu	Thr	Tle	Arg	·
	:		89								_								~~9	<b></b> .
GCC	בא			GAA		1098 דביד		. ь.	11 Tr c	.07	CCA	<b>TY</b>	11.	16 ∽ ₁	· ·	~~ ¹	125	~	1	134
								-												
ντσ	Va.	ιG	iy (	Glu	dev.	Ser	Pho	I.	le L	eù l	Gly	Cy:	• G	ly S	er l	Pro	Leu	Leu	Pro	ala
_		11				1152			11	61			117	70		1	179		1	198
CIC	ထ	\ T	GC (	STC.	CVC	$\infty$	ATC	A	33 Y	TA (	3GA	3	C	NC X	cr c	200	ccc	TTC	ञ्ळा	CCA
Val	er,	, C	ys I	ا م∨	Λ== 	Clv	Her	A:	T	le (	31v	Pro		no T	 117 A		 D-0	 Dh.	TIP (	~1
	•	-	-			,			-		,			، رو.	,	u	10	rnc	TID (	oth

Figure 10 (Continued)

# Thermotoga maritima Alpha-qalactusidano Complete Gene Sequenca (% 51%)

1197 1206 1215 1224 1233 1242 GAA CAT ATA GAA GAC AAC GGA GGT GT GGT GGT AG
THE
and his the Giu Asp Asn Cly Ala Pro Ala Ala Ary Trp Ala Lou Ary Asn Ala
The same and the s
Ally Tyr Phe Hat His Asp Arg Phe Trp Leu Asn Asp Pro 100
ATA CTG AGA GAG GAG AAA ACG GAT CTC ACA CAG AAG GAA AAG GAG CTC TAC TCG
Ile Leu Arg Glu Glu Lys Thr Asp Leu Thr Gln Lys Glu Lys Glu Leu Tyr Ser
TAC ACG TOT GGA GTG CTC GAC AAC ATC ATA GAA AGC GAT GAT CTC TCG CTC
Tyr The Cys Cly Val Leu Asp Asn Met Ile Ile Glu Ser Asp Asp Leu Ser Leu
1458 CAT CAT CEA ANA AND GIT CTC ANA GAN ACC CTC CEN CTC CTC
Val Ary Asp His Gly Lys Lys Val Leu Lys Glu Thr Leu Glu Leu Leu Gly Gly
THE TEN GAT CIG AGA TAC GAG ATTA
The Ary Val Gin Asn Ile Met Ser Glu Asp Leu Ary Tyr Glu Ile Val Ser
THE THE ACC STO AND ATC GTG GTC GAT CTG AAC ACC ACC
Ser Gly Thr Leu Ser Gly Asn Val Lys Ile Val Val Asp Lie And Car Ling Glu
TAC CAC CTG GAA AAA GAA GGA AAG TCC TCC CTG AAA AAA AGA GTC GTC AAA AGA
Tyr His Leu Glu Luc Glu Clu Clu
Tyr His Leu Glu Lys Glu Gly Lys Ser Ser Leu Lys Lys Arg Val Val Lys Arg
1629 1638 1647 1656 1665  TAN CAC GCA AGA AAC TTC TAC TTC TAC GAA CAG GCT GAG AGA GAA TGA 3
slu Asp Gly Arg Asn Phe Tyr Phe Tyr Clu Glu Gly Glu Arg Glu
A STR OTH GIA CIN YLA CIN ***

Figure 10c(Continued)

# Thermotoga maritima β-mannanase (saper) (66P.2)

											-								
			9			18			27		•				45			54	
5'	ATG	GGG	ATT	CCT	GGC	CYC	GAC	TCC	TGG	AGC	CCG	TCA	GTA	TCG	CCC	Gλλ	TTC	CLL	
	Met	Gly	Ile	Gly	Cly	α2Α	λsp	Ser	Trp	Ser	Pro	Ser	Val	Ser	Ala	Glu	Phe	Leu	
			63			72			81			90		•	99				
	TTA	تكلمك		CTT	CAC.	. –	TOT	<del>-1</del>		CTC	للململة		AGT	G) C			<del>ст</del> -	108	
	Leu	Leu	Ile	Val	Glu	Leu	Ser	Phe	Val	Leu	Phe	Ala	Ser	Asp	Glu	Phe	Val	Lys	
														-				•	
			117		•	126			135			144			153			162	
	GTG	Gλλ	AAC	GGA	YYY	TIC	GCT	CTG	AAC	GGA	YYY	GAA	TTC	AGA	TTC	ATT	GGX	AGC	
													77-						
	Val	GIU	ASD	GTA	rys	Pne	YIZ	Leu	ASII	CTA	гÃя	GIU	Phe	Arg	Pne	TTE	GIA	Ser	
			171			180			189			198			207			216	
	λλC	AAC			λTG		TAC	AAG			GGA		ATA	GAC			CTG		
	λsn	Asn	Tyr	TYI	Met	His	Tyr	Lys	Ser	asa	Gly	Xet	Ile	yeb	Ser	Val	Leu	Glu	
												252			5.50				
		~~	225			234		110	243		ACA	252	TGG	CCT	261		CNC	270	
	AGT		AUA								,								
	Ser	Ala	λra	λερ	Met	Gly	Ile	Lys	Val	Leu	Arg	Ile	TIP.	Gly	Phe	Leu	Asp	Gly	
						_		_									_	_	
			279			288			297			306			315			324	
	GAG	AGI	TAC	TGC	УGУ	GYC	AAG	AAC	YCC	TAC	ATG	CAT	CCI	GAG	ccc	GGT	GIT	TTC	
	C3.			. ~						~~~	War	ui e	Pro	Gly	Pro		U-1	Dhe	
	GIU	Sei	171	Cys	ALG	עפא	Lys	W211			1100			914		Gly	741	2	
			333	ı		342			351			360	,		369	1		378	
	GGG	GTG	CCA	GAA	GGA	ATA	TCG	AAC	GCC	CAG	AGC	GGT	TIC	GAA	AGA	CTC	GAC	TAC	;
	Gly	Va]	Pro	) Glu	Gly	Ile	Ser	Asn	Ala	Gln	Ser	GJA	. bpe	G1v	·Arg	Leu	yab	Tyr	
			387	,		396			405			414	!		423			432	,
	202	الملت ا			GCG			CTC	-		AAA		GTC	אַדיג			GTG		
	Thr	Va)	Ala	Lys	Ala	Lys	Glu	Lev	Gly	Ile	Lys	Lev	. Val	. Ile	. Val	Leu	[aV ن	. Asn	ı
														*					
			441			450			459			466			477			486	
	AAC	TGC	G CA	CAC	TTC	: GGT	GCA	ATC	AAC	CAG	TAC	GIC	λGC	TGC	TIT	r GGJ	GG	ACC	
	\	·			Dhe		, G1v	. Her		67.	· ~~	- Val	LArg	·	Phe		, 6),	· The	
	WRI.		, ,,	, vař	, 1110	. 513	G.J		. ,		3.		5						
			49			504			- 513			522			533			540	
	CAT	CAC	C GA	GA1	י דדכ	TAC	: AGA	CAT	r GAC	AAC	ATC	LAA	(GA)	CY(	TAC	S AA	L AAC	TAC	:
											. <b></b> .				- `				
	His	Hi:	B AS	b yal	Phe	xXx	: Arg	, Ası	Gl:	ı Lyı	ı Ile	e Ly	s Glu	ı Glı	ı Ty:	r Ly:	Ly:	з Туг	:

Figure 11a.

		÷		•	
Thermot	oga maritim	a β-mannanas	• ( <del>1200)</del> (	continued) (	らりよ
549	558	567	576	F 0.5	_
GTC TCC TTT C	TC GTA AAC CA	T GTC AAT ACC	TAC ACC CC1 ~	585	594
			THE NEW GON G	IT CCT TAC AGG	GAA
Val Ser Phe Le	eu Val Asn Hi	Val Asn Thr	Tyr The Cly V	1 7	
· ·			-, dry v	at Pro Tyr Arg	Glu
603	612	621	630	630	
GAG CCC ACC A1	nc are see to	GAG CTT GCA	AAC GAA CCG C	20 TC	648
				~ 101 GYG YCC	GAC :
Glu Pro Thr Il	e Met Ala Tr	Glu Leu Ala	Asn Glu Pro As	FG (A 0)	
	•		, , , , , , , , , , , , , , , , , , ,	A cas cin the	yab
657	666	675	684	607	
AAA TCG GGG AA	IC ACC CTC GT	CAG TGG GTG	AAG GAG ATG AC	20 ACC W/O ***	702
				TO THE ATA	AAG
Lys Ser Gly As	n Thr Leu Val	Glu Trp Val	Lys Glu Met Se	T Ser The II.	
•				" per tår lie	rys
711	720	729	738	747	756
AGT CTG GAT CC	C YYC CYC CIC	GIG GCI GIG (	SCG GAC GAA GC	A TIC TIC ACC	730
C *					AAC .
Ser Leu Asp Pr	o Asn His Lev	Val Ala Val (	Sly Asp Glu Gl	V Phe Phe Ser	1
					ഹം
765	774	783	792	801	810
TAC GAA GGA TT	C AAA CCT TAC	CET COL CALA	ECC GAG TGG GC	C TAC AAC GGC	TCG
Tyr Glu Gly Ph	e Lys Pro Tyr	Gly Cly Clu ;	la Glu Trp Al	a Tyr Asn Glv	Tro
819				-	
	828	837	846	855	864
TCC GGT GTT GA	C 100 AAG AAG	CIC CIT TCG A	TA GAG ACG GT	G CAC TTC CCC	ACG
Ser Gly Val As	b ith pas pas	red red Ser I	le Glu Thr Va	l Asp Phe Gly	Thr
873	882	891			
	ד ככפ זכנ ראר ד	3.00 CC2 C22 C22 7	900	909	918
TTC CAC CTC TA		TOG OGT GTC A	GT CCA GAG AA	C TAT GCC CAG	TGG
Phe His Leu Ty	r Pro Ser His	Trn Gly Val S			
		P OLY VAL 3	er Fro Giu As	n Tyr Ala Gln	Trp
927	936	945	954		•
GGA GCG AAG TGC	ATA GAA GAC	CAC ATA AAG A	TC GCA AAA GA	963	972
Gly Ala Lys Tr	Ile Glu Asp	His Ile Lvs I	le Ala fam Cl		
	-		was Dys GI	Tie Civ PAR	Pro
981	990	999	1008	1017	
GTT GTT CTG GA	GAA TAT GGA	ATT CCA AAG A	GT GCG CCA GT	1017	026
Val Val Leu Glu	Glu Tyr Gly	Ile Pro Lys S	er Ala Pro Va	l Aco Aco The	
				MI AIG INF /	#T¶
1035	1044	1053	1062	1071 11	080
ATC TAC AGA CTC	TGG AAC GAT	CTG GTC TAC G	AT CTC GGT GG	A GAT GOA GOG!	ATG
Ile Tyr Arg Leu	dev usy day	Leu Val Tyr A	p Leu Gly Gly	Asp Cly Ala I	for

Figure 11b(Continued)

		The	X30	toga	ma	riti	-	β- <b>3</b> 2	anna	Dasa		· (全)	٠.			1		0
	•	10	9 o								٠, ٢	MACE /	( c	onti	Dega	1) (		Γ,
TT	C TO	E A	TG C	TC (2	10 CG C	98 C) )	<b>-</b>	11	107	•	11:	16		1129	5		122	
							1C (	GG G	iaa G	GT T	CG CI	 16 YC	A GA	C GAC	S AGA	, éc	יכנג ימים	
Ph	e Tr	P M	et L	eu A	la G	ly I	le G	ly G	lu G	 lv s		op Ar						
		11	,,			_		•	•	-, 50	12	SP AF	g ye	P Glu	. Arg	C17	בעד ז	;
TA	T CC	G G	AC TO	AC C1	11!	52		11	61		117	0		1179				
					·		, C V	CA X	TA G	LC YY	C CA	C GA	C AG	ר ככא	GAA	GC	1188	
Ty	r Pr	o λ:	T)	/T λs	p GI	Ly Pr	le A	ra T	 le V:			C GAC  P Asi		·				
								-				p As	Ser	Pro	Glu	Ala	Glu	
CTC	3 AT	Y YC	)7 :a c:	l A ma	120	)6 		12	15		122	4		1233				
								16 1	IC A	C YC	λGG	T GAN	GAC	ATA	λGλ	GAA	1242 CAC	
Lev	ı Ile	s Ar	g G1	u Ty	T A1	a Ly	3 L	eu Pi	00 As			r GAA  Y Glu						
		126														Glu	λsp	
ACC	TGO	125 TC	لمدل مثل T	~ .	126	0		126	59	•	127	В.		1287				
						T CC	A A.	v. w	rë ee	C AT	G GA(	3 ATC	AAA	AAG	ACC	GIG	Cy 7	
Thr	ς.Σ.≅	Se	r Ph	e Il	e Le	u Pr	o Lv	'S As	n G1			l Ile						
		3 3 0	_	•										Lys	Thr	Val	Glu	
GTG	λGG	T30	5 r c:-	سن س	131	4 .		132	3		1332	2		1341		-	350	
						- GA	- TA	C YC	CAN	C YCC	TT	CAN 1	λAG	IIG	TCT	ണ്	AAA	
Val	Arg	Ala	Gl;	y Val	l Phe	: <b>λ</b> 5]	Ty	r Se	r Ası		·	Glu						
		1359					_				FILE	GIU	гЛЗ	Leu	Ser	Val	Lys	
GTC	GAA	CAT	י רכיזע	3 CTT	1368	3 7 Cli		137	7		1386		3	.395		,	404	
									a WIN	L GAG	CAT	, CIC	GGA	TAC	CC3	λΤΤ	TAC	
Val	Clu	Asp	Lev	ı Val	Phe	Glu	λει	n Gli	ı Ile	Glu	His	Leu		·				
	1	1273			1 1 2 2													
GGC	TIT	GAT	CTC	: GAC	1422	ACC	~	1431	l 		1440	GλΑ	1	449		1	45B	
							~			GAT	CCA	GλA	CAT	GAA ;	ATG 7	TTC	CIT	
CIA	Phe	Asp	Leu	(RED	Thr	Thr	Arg	Ile	Pro	Asp	Gly	Glu	Hie	 Clu v				
	1	467			3476													
GAA	GGC	CAC	TIT	CAG	GGA	AAA	ACG	1485 270	i. Lara		1494	ATC .	1	503		15	12	
										GAC	TCT	ATC	XXX (	GCG A	LAA G	TG (	TG	
Glu	CIA	HIS	Phe	Gln	Gly	Lys	The	Val	Lys	λsp	Ser	Ile	Lvs	 Alm :	 V			
	- 1	571		•														
AAC (	Gaa 🔻	GCA	CGG	TAC	GTG	CTC	GCA	GAG	GAA	مصنت آ	.548		1	557		15	66	
											CAT	TIT	rcc 7	CI C	CA G	AA G	AC	
neA	alu ,	WT 9	Arg	Tyr	Val	Leu	λla	Glu	Glu	Val	Αsp	Phe S	Ser S	er Þ		 }		
	7 1	575		1	E D 4													
GTG ;	LAA J	AAC	TGG	TGG	AAC	AGC	GGA	224	TGG	CAG :	602 CCN	C\ C -	16	11		16	20	
 Val 1	·											UAG 7	rrc G	GG I	CY C	CT G	λC	
Val 1	·ys j	'SU	TTD	Trp	Asn	Ser	Gly	The	Trp	Gln ,	Ala	Glu F	he G	ly s	 er D			
														_, _,			-V	

Figure 110 (Continued)

												-	٠.					
		The	mot	OGR	<b>3</b> 43	riti		B- <b>na</b> :	ppan	220	Œ	<b>3</b>	( c	:onti	Due	<b>a</b> ) (	(66	د م
AT	T CA	162	29		16-												1674 A CTG	•
11	- G1	u Tr	P λs	 m G]	 ly G1	ບ Va	 il ci	 v le						G AA		3 AA 	λ CTG  s Leu	
		TOR	3		160	1												
ccc	: cc	A AA	G AG	ငယ	C 1G	G GX	УGУ	170 A GT	1 G AG	A GT	171 'A GC	0 A AG	G AA	1719			1728 A CTC	
Pro	G1	y Ly	s Se	 × λs	p Tr	p G1	 u G1	 u Va	1 Ar	 7 Va			Tae		- GA	· AG	A CTC  I Leu	
			,		174	c												
	GA		r ca	G AT 	 C CI	C GA	G TA	C CY(	TA :	TA	ב אז	r cci	A AAC	5 GJC	GAG	GGJ	1782 CTC	
Ser	Gl	ı cy:	s Gl	u Il	e Lei	u G1:	נגנ ה	. yzi	Ile	Ty	r Ile	e Pro	Asr	Val	Glu	Gly	Leu	
		1791	L		180	n.											1836 GGC	
Lys	Gly	· Arg	Let	 1 Ax9								: GGC	TGG	GTG	AAG	ATA	GGC	
		1845	,		1054	1												
CIC	GAC	ATG	AAC	AAC	GCG	AAC	GT.	GAN	AGT	GCG	1872 Gag	ATC	ATC	1881 ACT	TTC	C.C.	1890	
Leu	Asp	Met	Asz	Ası	Ala	λsn	Val	Glu	Ser	Ala	Glu	Ile	 Ile	Thr	 Phe	Glv	Gly	
		1899			1900													
 Lvs						CAT	GTA	AGA	ATT	CYC	TTC	GAC	AGA	1935 ACA	GCG	ဇဇေ	GTG	
			Arg	Arg	PDe	His	Val	λrg	Ile	Glu	Phe	Asp	Arg	Thr	Ala	Gly	Val	
AAA	CAA.	7223	~`~		1962		-	1971		:	1980			020		-		
			CAC	ATA	GGA	GTT	CIC	CCI	CAT	CAT	CIG	AGG	TAC	1989 GAT	CCA	ردن 1	ያያይ	
Lys	Glu	Leu	His	Ile	Gly	Val	Val	Gly	qeK	 His	Leu	 Arg	Tyr	Asp	Glv	 Pro	71a	
		2007			2016	•	-										-75	
TTC	ATC	GAT	AAT	erc.	AGA	CTT	TAT	220	202	200	2034		2	043				
										ACA	GGA	GCT	ATG	TGA	3•			
Phe :	Ile	Asp	ysu	Val	Arg	Leu	Tyr	Lys	λrg	Thr	Gly	Glv	Mer	•••				

Figure 11d (Continued)

### APPII la β-mannosidase (63GB1)

. 9	18			
5' ATG CTA CCA GAA GAG	TIC CTA TCC	27 3	6 4	5 ₅
		CCC CTT CCC CA	C TCA GGC TT	T CAG TTC GA
Met Leu Pro Glu Glu	Phe Leu Trp	Gly Val Gly Gly		
ATG GGC GAC AAG CTC	72	81 9	D 9,	
ATG GGC GAC AAG CTC	NOS NGG CAC	ATC GAT CCA AA!	F ACC GAC TGG	TGG AAC TO
Met Gly Asp Lys Leu	Arg Arg His	Ile Arm Dec >		100
Met Gly Asp Lys Leu		-re wab blo Var	The Asp Trp	TIP LYS TIE
GTT CGC GAT CCT TTC	UC ATA AAA J	VAG GAG CIT GIG	ACT GGG CAC	162
Val Arg Asp Pro Phe	en Tle Inc.			CIT CCC GAG
	TIG LYS !	As Cln Len Asl	Ser Gly Asp	Leu Pro Glu
1/1	DA			
GAC GGC ATC AAC AAC	AC GAA CIT I	TT GAA AAC GAT	207 CAC AAG CTC	216
Asp Gly Ile Asn Asn Asn				GCT AAA GGC
Asp Gly Ile Asn Asn T	At CIH Len b	he Glu Asn Asp	His Lys Leu	Ala Lvs Glv
275	<b>7</b> /			
CTT GGA CTC AAC GCA T	AC AGG ATT G	GA ATA GAG TGG	261	270
Leu Gly Leu See 22.			NOC AGA ATC	MA CCC LCC
Leu Gly Leu Asn Ala T	Ar yrd Ile C	ly Ile Glu Trp	Ser Arg Ile	Phe Pro T-
2/9 2				
CCG ACG TGG ACG GTC G	AT ACC GAG G	306 20 377 369 27	315	324
Pro The The Table 1			ACT TAC GGT	TTA GTA AAG
Pro Thr Trp Thr Val As	p Thr Glu Va	l Glu Phe Asp	Thr Tyr Clv	Leu Val Tan
. 333 9/				
CAC GTT AAG ATA GAC AA	G TCC ACC CT	360	369	378
GAC GTT AAG ATA GAC AA		- GCI GVY CIG	GAC AGG CTG	GCC AAC AAG
Asp Val Lys Ile Asp Ly	s Ser Thr Le	u Ala Glu Leu	ASD ATT IOU	
387 30	~			•
GAG GAG GTA ATG TAC TA	40 AGG CGC CT	5 414	423	432
Cl		ATT CAG CAT	TTG AGG CAG (	TC GGC TTC
Glu Glu Val Met Tyr Ty	r Arg Arg Val	l Ile Gln His		
441 45			and wid GIR I	eu Gly Phe
AAG GTC TTC GTT AAC CTY	459	468	477	496
AAG GTC TTC GTT AAC CTC	- we far TAC TAC	ACG CTT CCA A	ATA TGG CTC C	AC GAC CCG
Lys Val Phe Val Asn Les	Asn His Phe	Thr Leu bes		
495 504	3	seu Pro I	Te Trp Leu H	is Asp Pro
ATA GTG GCA AGG GAG AAG	513	522	531	E40
ATA GTG GCA AGG GAG AAG	CCC CTC ACA	ANC GAC AGA A	TC GGC TGG G	540 TC TCC CAG
Ile Val Ala Arg Glu Lys	Ala Leu The	A==	<u></u>	
-		wen wab wil I	le Gly Trp V	al Ser Gln

Figure 120

	-	•	<b>LEP</b> I	I 1	■ B-	man	teod	dase	• (	63 G B	1)	(co:	atin	(ber	)		
		549			558	1		5.67			F 7 6						
AGG	ACA	GTI	CII	. CYC	TTT	GCC	AAG	TAT	CCT	GCT	576	3.770		585			594 GGA
λrg	Thr	. Val	Val	Glu	Phe	Ala	Lys	Tyr	λla	Ala	Tyr	Ile	Ala	His	Ala	Leu	Gly
		603			612			621			630			639			
GAC	CTC	GTG	CYC	YCY	TGG	AGC	ACC	TTC	AAC	CAA	CCT	λTG	GTA	ىلىك دەت	تبلت	G) C	648 CTC
Asp	Leu	Val	Asp	Thr	TIP	Ser	Thr	Phe	λsn	Glu	Pro	Met	Val	Val	Val	Glu	Leu
		657			566			675									
GGC	TAC	CTC	GCC	ccc	TAC	TCA	GGA	طمامال ۱۳۰۵	~	CCG	603	~~~		693			702
Gly	TYE	Leu	Ala	Pro	lar	Ser	Gly	Phe	Pro	Pro	Gly	Val	Met	Asn	Pro	Glu	λla
•	•	711		•	720			729			738			747			
CCC	AAG	CIG	GCG	ATC	CTC	AAC	ATG	ATA	AAC	CCC	CAC	GCC	TTG	GCA	ጥልጥ		756
VIF	Lys	Leu	Ala	Ile	Leu	Asn	Met	Il.	λsn	Ala	His	λla	Leu	λla	Ţyr	Lys	Met
		765			774			783			792			801			810
ATA	λλG	AGG	TTC	GAC	ACC	AAG	AAG	GCC	GAT	GAG	GAT	AGC	λλG	TCC	CCT	CCC	GYC
									~~-				_				
178	Lys	<b>λ</b> rg	Phe	ysb	Thr	Lys	Lys	λla	yzb	Glu	уар	Ser	Lys	Ser	Pro	ملم	Asp
		819			828			837			846			055			
CTT	GGC	AΤλ	ATT	TAC	AAC	AAC	ATC	GGT	GIT	GCC	TAC	CCT	222	855	~~		864
Val	GΙΆ	Ile	Ile	ÍΣ	Asn	Asn	Ile	CJA	Val	Ala	Tyr	Pro	Lys	Asp	Pro	λsn	λsp
		873			882			891			900			909			918
CCC	AAG	CAC	GIT	λλλ	GCA	CCC	GAX	AAC	GAC	AAC	TAC	TTC	CAC	AGC	GGA	CTG	TTC
	<b>-</b>	Asp		Lys					Asp	Asn	Tyr	Phe	His	Ser	Gly	Leu	Phe
بلملمك	C > T	927			936			945			954			963			972
	GAT	GCC	ATC	CAC	AAG	CCT	AAG	CTC	AAC	ATA	GAG	TIC	CYC	GGC	GAA	AAC	TIT
		λla		()		Gry	rys	red				Phe	Asp	Gly	Glu	λsn	Phe
CT'A		981		~~	990			999		3	1008		3	1017		1	.026
		CTT	<b>AUA</b>	CAC	CTA	AAA	GGC	AAT	GAC	TGG	λTλ	GGC	CTC	AAC	TAC	TAC	ACC
Val	Lys	Val	Arg	His	Leu	Lvs	Glv	Asn	Asn								
		1035	_									GIY	Pen	ABD	TYT	TYT	Thr
CGC			GTT		.044 TAT	ייטע	Cac I	.053		3	062		1	.071		1	080
		CTT				1			AAG	TIC	CCY	AGT	ATA	ccc	CTC	ATA	TCC
Arg	Glu	Val	Val	λrg	Tyr	Ser	Glu	Pro	Lvs	Phe	Pro	Ser		D===			
				-	-	_			_, .			-41	T T E	rro	rea	110	>81

Figure 12b(Continued)

## p-mannosidase (630B1) (continued)

1000
1089 1098 1107 1116 1125 113
AAG GGC GTT CCC AAC TAC GGC TAC TAC TAC TAC TAC TAC TAC TAC TAC TA
TTC AAG GGC GTT CCC AAC TAC GGC TAC TCC TGC AGG CCC GGC ACG ACC TCC GC
Phe Lys Gly Val Pro Asp To Gly
Tyr Gry Tyr Ser Cys Arg Pro Gly The
Phe Lys Gly Val Pro Asn Tyr Gly Tyr Ser Cys Arg Pro Gly Thr Thr Ser Al
GAT GCC ATG CCC GTC AGC GAT ATC GGC TGG GAA GTC TAT CCC CAG GGA ATC TAG
1179 118
THE TAX GIVE TAX CCC CAG GGA ATC TAG
Asp Gly Met Pro Val Ser Asp Ile Chy man
Asp Gly Met Pro Val Ser Asp Ile Gly Trp Glu Val Tyr Pro Gln Gly Ile Tyr
1197 1206 1215 1224 1233 1242
GAC TCG ATA GTC GAG GCC ACC AAG TAC AGT GTT CCT GTT TAC GTC ACC GAG AAC
ARD SOUTH AND GIT ACC GAG AAC
Jest lie val Glu Ala Thr Lys Tyl Ser Val Des
Asp Ser Ile Val Glu Ala Thr Lys Tyr Ser Val Pro Val Tyr Val Thr Glu Asn
1251 1260 1269 1278 1287 1296 GGT GTT GCG GAT TCC GCG GAC ACG CTG AGG CCA TAC TAC AGG CTA TAC TAC TAC TAC TAC TAC TAC TAC TAC
GGT GTT GCG GAT TCC GCG GAC ACG CTG AGG CCA TAC TAC ATA GTC AGC CAC GTC
Gly Val Ala Asp Car Mar
Ata Asp Thr Leu Arg Pro Tyr Tyr Ile Val
Gly Val Ala Asp Ser Ala Asp Thr Leu Arg Pro Tyr Tyr Ile Val Ser His Val
1305 1314 1323 1332 1341 1350
ATA GAG GAA GCC ATT GAG AAT GGA TAC GCC CTC 1341 1350
Ser Lys Ile Glu Glu Ala Tle Glu 1-
Ser Lys Ile Glu Glu Ala Ile Glu Asn Gly Tyr Pro Val Lys Gly Tyr Met Tyr
1359 1368 1377 1386 1395 1404
TGG GCG CTT ACC CAT 330 1377 1386 1395
1404
TITE ALE LEW The Act of the Control of the Act of the A
Trp Ala Leu Thr Asp Asn Tyr Glu Trp Ala Leu Gly Phe Ser Met Arg Phe Gly
The Ser Met Arg Phe Gly
CTC TAC AAG GTC GAC CTC ATC TCC AAG GAC ACC 100 1449 1458
THE WAS AGG ATC CCC ACC ACC
Leu Tyr Lys Val Asp Leu Ile Ser Lys Glu Arg Ile Pro Arg Glu Arg Ser Val
Ile Ser Lys Glu Arg Ile Pro Arg Glu Arg
1467 1436
GAG ATA TAT COR 100 14/6 1485 1494
1467 1476 1485 1494 1503 1512 GAG ATA TAT CGC AGG ATA GTG CAG TCC AAC GGT GTT CCT AAG GAT ATC AAA GAG Glu Ile Tot Acc
THE TYP Arg Arg Ile Val Gln Ser has Glass
Glu Ile Tyr Arg Arg Ile Val Gln Ser Asn Gly Val Pro Lys Asp Ile Lys Glu
GAG TTC CTG AAG GGT GAG GAG AAA TGA 3
THE CALL DATE AND TOTAL 3.
Glu Phe Leu Lys Gly Glu Glu Lys ***

Figure 12C(Continued)

### OCI/4V Endoglucanase (33GP)

	9	18						
5' ATG GTA G	M AGA CA	C TTC A	CA TAT	27		36	CTG TTT CTT	5.4
					T ATT T	CC YCC	CLG LLL CLL	GTT ATC
met Val Gi	u Arg Hi	s. Phe Ai	rg Tyr	Val Le	u Ile C		Leu Phe Leu	
6	3					75 III	Leu Phe Leu	Val Met
TA KTO OTO	C TCA TC	72 C ) (77 ) (1		81		90	99	•••
			IG TGT	GCY YY	A AAT G	AA CCA	99 AAC AAA AGA	20E
Leu Leu Il	e Ser Se	Thr G1	n Cvs	Gly Lan				
				, 2,	· Nan G	u Pro	Asn Lys Arg	Val Asn
		A GTT GC	T GAA	yel cyi	YEC Y	C TCA	153 GCA TTT GAA	162
Ser Met Gl	u Gln Sex	Val Al						TAC AAC
•			a did	ser Asp	Ser As	m Ser ]	Ala Phe Glu	Tyr Asn
17	1	3 0 0						
AAA AIG GI)	A GGT AAA	GGA GT	A AAT A	ATT CGA	AAT GO	T TTA (	207 SAA GCT CCT	216
Lys Met Val	Gly Ive	Class 12-					and GCT CCT	TIC GAA
	· vij bys	GIY VE.	r war :	Ile Gly	yzu YJ	a Leu G	Slu Ala Pro	Phe Glu
225	i	274	_				*	
GGA GCT TGG	GGA GTA	AGA ATT	י פאפ ל	SAT GAA	TAT TT	ፈ ፓርኳር አ	261 ATA ATA AAG	270
Glv Ala Tom			·				TA ATA AAG	YYY YCC
Gly Ala Trp	GIA AST	Arg Ile	Glu A	rab CJ n	Tyr Ph	e Glu I	le Ile Lva	Lare hee
279		200	_					
GGA TTT GAT	ACL CLI	AGG ATT	, CCC y	TA ACA	300	5	315	324
Class Division						GCA C	AT ATA TOO	EAX AAG
Gly Phe Asp	Ser Val	Arg Ile	Pro I	le Arg	Trp Ser	Ala H	 is Tle c (	
333		343						
CCA CCA TAT	GAT ATT	GAC AGG	3 227 m	51	360	•	369	378
CCA CCA TAT				TO CIC	GYY YCX	GIT A	AC CAT GTT G	TC GAT
Pro Pro Tyr	Asp Ile	Asp Arg	Asn P	he Leu	Glu Arc	Val 2		
387						, m ± 7°	an HIS AWI A	al Asp
AGG GCT CTT	GAG AAT	395	303	05	414		423	432
AGG GCT CTT			ACA G	IN ATC	ATC AAT	YCC C	C CAT TIT G	AA GAA
Arg Ala Leu	Glu Asn	Asn Leu	The Va	al Ile	 Tle Acn			
443					-16 7311	INE HI	s His Phe G	lu Glu
CTC TAT CAR	CAN CCC	450	45	9	468		477	406
CTC TAT CAA		GAT AAA	TAC GG	C GAT (	SIT TIG	GTG GA	A ATT TEG A	SA CAG
Leu Tyr Gln	Glu Pro	Asp Lvs	Tvr 61	·				
Leu Tyr Gln		,5	-31 01	y ASD \	/al Leu	Val Gl	u Ile Trp A	g Gln
495		E A 4						
ATT GCA AAA	TIC TIT 1	WA GAT	TAC CC	C GAA A	AT CTG	TTC TT	7 GAA 3~~ ~-	540
Ile Ala Lva	Phe Phe	ve les					- OAR AIC T	IC AAC
Ile Ala Lys :		AR V2D	IYI PI	o Glu A	sn Leu	Phe Ph	e Glu Ile T	'I Asn

Figure 13a

OC1/4V Endoglycanas (Ann.
OC1/4V Endoglucanase (33GP1) (continued) 549 558 567 576 585
GAG CET COT CAG AND
GAG CCT GCT CAG AAC TTG ACA GCT GAA AAA TGG AAC GCA CTT TAT CCA AAA GTG Glu Pro Ala Gln Asn Leu Thr Ala Glu Lys Trp Asn Ala
Glu Pro Ala Gln Asn Leu Thr Ala Glu Lys Trp Asn Ala Leu Tyr Pro Lys Val
And Gid Lys Trp Asn Ala Leu Tyr Pro Lys Wall
603 612 633
CTC AAA GTT ATC AGG GAG AGC AAT CCA ACC CGG ATT GTC ATT ATC GAT GCT CCA
THE TATE CON ACC CGG ATT GTC ATT ATC GAT GCT
Leu Lys Val Ile Arg Glu Ser Asn Pro Thr Arg Ile Val
Leu Lys Val Ile Arg Glu Ser Asn Pro Thr Arg Ile Val Ile Ile Asp Ala Pro
657 666 675 694
AAC TGG GCA CAC TAT AGC GCA GTG AGA AGT CTA AAA TTA GTC AAC GAC AAA CGC
AGA AGT CTA AAA TTA GTC AAC GAC AAA CCC
Asn Trp Ala His Tyr Ser Ala Val Are Ser
Asn Trp Ala His Tyr Ser Ala Val Arg Ser Leu Lys Leu Val Asn Asp Lys Arg
ATC ATT GTT TCC TTC CAT TAC TAC GAA CCT TTC AAA TTC ACA CAT CAG GGT GCC
THE ANA TTO ACA CAT CAG GGT GCC
Ile Ile Val Ser Phe His Tyr Tyr Glu Pro Phe Ive Pho Chi
The The His Gln Glv 11a
THE GIT AAT CCC ATC CCA CCT GIT AGG GIT AAG TOTAL BEIL 810
GAA TGG GTT AAT CCC ATC CCA CCT GTT AGG GTT AAG TGG AAT GGC GAG GAA TGG
Glu Trp Val Asn Pro Ile Pro Pro Val Arg Val Lys Trp Asn Gly Glu Glu Trp
GAA ATT AAC CAA ATC AGA AGT CAT TTC AAA TAC GTG AGT GAC TGG GCA AAG CAA
BEA AND AGA AGT CAT TTC AAA TAC GTG AGT GAC TGG CGL LIGHT
Glu Ile Asn Gln Il
Glu Ile Asn Gln Ile Arg Ser His Phe Lys Tyr Val Ser Asp Trp Ala Lys Gln
AAT AAC GTA CCA ATC TTT CTT GGT GAA TTC CCT GGT 909 918
THE GOT GET TAT TO ARE
Asn Asn Val Pro Ile Phe Leu Gly Glu Phe Gly Ala Tyr Ser Lys Ala Asp Het
of the Gly Ala Tyr Ser Lys Ala Asp Met
GAC TCA AGG GTT AAG TGG ACC GAA AGT GTG AGA AAA 963 972
GAC TCA AGG GTT AAG TGG ACC GAA AGT GTG AGA AAA ATG GCG GAA GAA TTT GGA
ASP Ser Arg Val Lys Trp Thr Glu Ser Val Arg Lys Mon 1
Asp Ser Arg Val Lys Trp Thr Glu Ser Val Arg Lys Met Ala Glu Glu Phe Gly
TIT TCA TAC GCG TAT TGG GAA TIT TGT GCA GGA TIT GGC ATA TAC GAT AGA TGG
Phe Ser Tyr Ala Tyr To Clark
Phe Ser Tyr Ala Tyr Trp Glu Phe Cys Ala Gly Phe Gly Ile Tyr Asp Arg Trp
TCT CAA AAC TGG ATC GAA CCA TTG GCA ACA GCT GTG GTT GGC ACA GGC AAA GAG
THE GCA ACA GCT GTG GTT GGG ACA
Ser Gln Asn Trp Ile Glu Pro Leu Ala Thr Ala Val Val
The Gly The Gly Type Cly
TAA 3.
***

Figure 136(Continued)

### Thermotoga maritima Pullulanase (6023)

					9			1	8			27											
:	, Y	rG	GΆ	T C	TT	ACJ	AA	G	G GC	G A	70	AT.	C77	~		36			•	5			54
													GIV	- A(	<b>عن ر</b>	TG	XXC	: GA	3 70	S C	AG (	GCA	54 333
	Me	e t	λs	p L	eu	Thr	Ly	s Va	1 G	v 'T	م۱	 Tla	37-										
							-			., _	16	T 7 G	Va.	LA	g L	eu	nak	Gli	a Tr	P G	ln i	Ala	Lys
					63			~	~					•									
	GA	C	CIV	G	CA	λλλ	GA	. AG	 G 7-1	~ ·	TR .	81				90			9	9.			108 TGG
		-								_ A.	TV (	فكالمد	ATA	r y	N C	ХC	CCY	AAC	GC	TG	u c	:77	100
	λs	P	Va:	LA	la	LVs	Asr		⇒ Dh												-		TEP
							,	, ,,,	9 -	e 1.	re (	ilu	Ile	: Ly	'9 A	S P	Cly	Lys	: A1	a Gi	ա ւ	- 7a 1	~
				1	17			3.3	_														
	AT	λ	CTO	: c	AG	GGA	CTC		. ~	<b>-</b>	1	135			1	44			15.	3			162 AGA
		_;					~	, GA	s GA	G A	L 1	TC	TAC	GA	y y	NA.	CCA	GAC	λC	- A .TY	~ ~		102
	Il	8	Lev	G	מו	G) v	V=1																NUA
				•		Gry	ATT	GII	1 G1.	u Il	e P	,pe	TYI	G1	u L	28	Pro	Aso	Th	- 50	- D		Arg
				17	71	•	•															10	лгg
	ATY	٠ -	TTC	-	· ·	ccs	CNC	180	,		1	89			19	8			20.	7			27.5
					-	~~	CAG	GCA	, AG	TC	G A	УC	λλG	GI	G An	TC (	GAG	GCT	7-1-	, ,	c ,	~	216
	I1	<b>e</b> 1	Phe	PH	10		C1-									<b>-</b> .					ол 		AAT
						<u> </u>	Gln	VID	AT	7 Se	I A	SU	Lys	Va:	1 11		Glu	Ala	Phe	م. ا			· · · ·
				22	5																<b>u</b> 1.		ASD
	ces		T	63		200		234			2	43			25	2			261				770
						n	λλλ 	AAG	AA	CA	УC	TC	TIC	AAC	G GI	T	CT	GTT	GAC	. C.	<b>x</b> x		270
	Pro	, ,	/al	λ «	- ·	Ph-	7															~~	وللمن
					,	• • • • •	Lys	rys	Lys	G11	u L	eu	Pbe	Lys	Va	1 1	Thr	Val	Asp	G)	 - T.		
				27	•															<b>U</b> 1.	, 1	/5	CIA
	ATT	· c	cc	<u></u>	, ,	~~	300	288			25	97			30	6	•		315				~~.
		_					ÀGX 	GIG	GAA	AAC	G GC	CC (	CAT	CCC	: AC	G	iAC .	ATA	GAC	CTY		<u> </u>	324
	Ile	P	70	V»	7 0	:	<b></b>													010	, ,,	ر ق	RAC
		_		• •		,eı	λrg	AFT	Glu	Lys	נגיי	La J	ds/	Pro	Th	<b>-</b> >	Sp.	Ila	Asn	Val	- TD		
				33:	3												-			V 44.2	11		ren
	TAC	G	TG	AG:		₩.	~~~	342			35	1			36	0			369				
		_		7.00		.10	erc	CIT	TCT	CYY	TC	25	TG	λλλ	GA	۱ G	<b>AA</b> (	GAC	<u>م</u>	AC N			3/8
	Tyr	·v	a l	1-		1-														NOA		A (	AC
	-	-		;	<i>,</i>	TA	Val	Leu	Ser	Glu	. 5e	r L	æц	Lys	Gli	ı G	luj	az.	I.en	7		- :	
				387	,															ar y	ւ	8 ,	rzb
	GTG	G	A.A	CTC:		· ·		396			40	5			414	l			423				
						,	ATA	GAA	GGT	TAC	λλ	y C	CG	GCA	λGZ	G	rc ,	TC.	ATG	. TY		- 4	32
	Val	G	Lu	T.	T .	 la 1	·									-				~10	(34)	5 A	110
	Val				-	16 7	rie (	31 <b>u</b>	Gly	TYI	Ly	5 P	ro 1	Mla	Arg	V	al I	le i	Mer	Mos			
				441																416 C	GI	4 1	16
	CTG	G			Tro			150			45	9	·		468				477				۰
	CTG		-		- 17	1C 1	AT	rac	GAT	GGA	GA	3 6	TC (	GA	GCC	G	<b>Γ</b> λ τ	'אר'	4-7T	CC>	~		86
	Leu	λ.	•		~															-	GA	> A	AG
	Leu		ر س	-up	1.7	T T	AT 1	JI.	yab	Gly	Gli	ש נ	eu C	ijу	λla	Va	11 7	ν <u>τ</u> '	Ser	D	<u></u>	:	
				195													- •	<i>_</i>		* T. O	01 l	L	y <b>s</b>
	ACG	ΑТ	י 1 A'			. ~	<b>-</b>	04			513	3			522				531			_	
	ACG				70	G	10 1	. GG 1	ICC	CCC	CIJ	יו יו	T A	λG	TCC	GT	'A A	AG (		محدث		· ~	40
	Thr	11	e -	 																-11	- 14	Т.	IC.
	Thr				A.	y V	di I	TP S	er	Pro	Val	Se	er L	УЗ	TIP	٥٧	1 L	vs 1	 /al	 ! e			
															-			'		₽EU	Let	P	1 <b>C</b>

Figure 14a

			The	) I'M	)tog		ari	time		wll:	alan		18		(20				
				349									( 0	<b>GF</b> 3)	( 00	onti	nued	i)	
	AAJ		) )	200	<b>~.</b> .		558			56	7		57	6					
						GAC ,	YCY	Gλλ	CCC	TA	CAC	G GT	r Gr	G AAC	` ATG	202	m		594
	Lvs	: A:	an G	1	G)	<b></b>										GAA	TAC	AAG	GCX
	-			-3	JIU ,	vzb .	Thr	Glu	Pro	נגנ ו	Gli	val	l Va	l Asn	Met	G3			
2	AAC	GC	≆G G	TC :	rcc (	7 7	512			621	•		630	)		639			
•							~ ·	GI-I	GIT	GΥY	, eec	: GY1	CIC	GAC	GGA	GTG	<b>TTY</b>	Th (	648
1	ne.	G	y V	al o	מבי	Slu J	110								Gly				Cic
								<b>V</b> 41	ATT	Glu	Gly	ysb	Leu	qeK i	Gly	Val	Phe	7~	Lev
7	TAT	CA	C C	rc c	ia ,	IAC T	AC (	GGA.	NG	675			684	ı	GAT	693			702
-										AIC	٨٤٨	ACA	ycc	GIC	GAT	CCT	TAT	TCG	AAA
7	λī	Gl	n L	eu G	lu A	Sn T	) IX	ilv	Lvs	Tle	\				Asp				
							_	-	-3-	+10	AL U	Inr	Thr	Val	yzb	Pro	Tyr	Ser	Lys
			- / 3	17		-	~ ~												
-		GT	T T	rc c	CY Y	AC A	AC C	:AA:	CAG	AGC	GCC	بلبك	120	3.3 m	CII	747			756
<b>A</b>	1-	7/-												~~1	CTP	GCC	ycc	УСУ	YYC
^		Va.	רו ד	∕⊈⊹ እ	la A	BD Y	en G	ln (	Glu	Ser	Ala	Val	Val	Asn	Leu	11-			
			76	e											Deu ,	ΛΙα	Arg	Thr	YZIJ
. с	C.A.	GAZ	, /o	ים ישיני	~ ~	7:	74			783			792			201			
_					~ G	AA AJ	VC G	AC I	AGG	GGA	೦೦೦	λλλ	ATC	Gλλ	GGA '	TAC	333 A		810
P:	ro	Gli	2 G1	V T:	e.	1 2.												JAC (	GCG
				•			10 A	вр	ra	Gly	Pro	Lys	Ile	Glu	Gly	Tyr (	Glu i	Asp :	
			81	9		27	0												
A:	CA.	ATC	TA	T G	נג גו	וא כא	C A	ra c	5CG 1	637 637	3700		846		GYY 1	855		1	864
											A1C	ACA	GGA	CTC	GAA )	WC :	ree d	GG (	FTA
11	le	Il.	TY:	r Gl	n II	e Ri	s I	le A	la	des	Ile	Thr.	G3 ve	·	Glu )		•		
												****	GIĀ	ren i	Glu )	מבו	er c	ily t	/al
	ъ.		87:			88	2		1	891			900						
			AA.	A GG	c cr	C TA	T CI	C C	GG (	ere i	ACC (	GAA (	GAA.	AAC :	e Acg a	עטי אא	·	9	18
Ly	·s )	Lsn	Lare														GA C	CC C	:GC
•			٠,٠	. 01	A re	u ry	r Le	u G	ly I	eu :	Chr (	Glu (	Glu /	Asn ?	hr L	VS G	lv b		1
			927	7		031	_												
GG	T	TG	ACA	AC	A GG	<del>مدن</del> ت	r •~	. c	9	45			954		9 A TE	63		. 9	72
									AC C	TT C	arg (	SAA C	בבכ (	GT (	TT A	CX C	AC G	TT C	λT
G1	y V	'al	Thr	Th	r Gl	y Lei	ı Se	- U.							al T				
					-			- 4.	ע בו	eu v	ar (	ilu I	eu (	ily V	al T	hr H	is V	al H	is
			981			000													
ATZ	A C	TT	CCI	TIC	TT	r GAT	TI	C TA	IC A	CA C	בר ר		801		10: AT A	17		10	26
						·				<del></del> -			AA C	TC G	AT A	AY C	T T	rc cu	AG
116	3 L	eu	Pro	Ph∈	: Phe	: Asp	Phe	e Ty	T	pr C	lv A	sp C	1,, 7		BD Ly				
			026					_		_	-,	- <b>-</b> - 0		au A	אס עם	/s A	P Pi	e GI	Lu
ኢኢር	· T	У. Т	035			1044	_		10	53		10	62		1.05	71			
			4AC	~^0	TGG	GGT	TAC	GA	TC	T T	AC C	TG T	TC A	TG C	107 IT CC	ים הי	·	108	10
Lys	T	<b>/</b> I	Tvr	lon	~												ان لان	K AG	.A
_		-		AB []	irp	GIA	TYI	λs	p Pi	(L 0.	YT L	eu P	he M	et V	al Pr	ים מי	" C		- -
							:									- 31	n 61	y AI	9

Figure 14b(Continued)

Thermotoga maritima Pullulanase (SGP3) (continued)
full timese (6GP3) (continued)
TAC TCA ACC GAT CCC AAA AAC CCA CAC ACG AGA ATC AGA GAA GTC AAA GAA ATG
Tyr Ser Thr Asp Pro Lys Asn Pro His Thr Arg Ile Arg Chart
Tyr Ser Thr Asp Pro Lys Asn Pro His Thr Arg Ile Arg Glu Val Lys Glu Met
1143 1152 1161
GTC AAA GCC CTT CAC AAA CAC COM 1170 1179
GTC AAA GCC CTT CAC AAA CAC GGT ATA GGT GTG ATT ATG GAC ATG GTG TTC CCT
Val Lys Ala Leu His Lys His Gly Ile Cly
Val Lys Ala Leu His Lys His Gly Ile Gly Val Ile Met Asp Met Val Phe Pro
1197 1206 1215 1224
TAC GG! ATA GGC CAA CTC TOT COR COR 1233
CAC ACC TAC GGT ATA GGC GAA CTC TCT GCG TTC GAT CAG ACG GTG CCG TAC TAC His Thr Tyr Gly Ile Gly Glu Leu Ser Ale Pho American
His Thr Tyr Gly Ile Gly Glu Leu Ser Ala Phe Asp Gln Thr Val Pro Tyr Tyr
1251 1260 1260
TTC TAC AGA ATC GAC AAG ACA COM COM 1278 1287 1287
The same and the s
Phe Tyr Arg Ile Asp Lys Thr Gly Ala Tyr Leu Asp Clura
Ser Gly Cys Gly Asp
GTC ATC CC3 300 01314 1323 1332
GTC ATC GCA AGC GAA AGA CCC ATG ATG AGA AAA TTC ATA GTC GAT ACC GTC ACC
Val Ile Ala Ser Glu Arg Pro Met Met Arg Lya Phe Ile Wal
The lie val Asp Thr Val Thr
1359 1368 1377 1386
1404
Tyr Trp Val Lys Glu Tyr His Ile Asp Gly Phe Arg Phe
THE ASD GIR MOT COLOR
1413 1422 1431 1440 1440
**** UNL AAI 130 101 140
ATC GAC AAA AAG ACA ATG CTC GAA GTC GAA AGA GCT CTT CAT AAA ATC GAT CCA  Lle Asp Lys Lys Thr Met Leu Glu Val Glu Arg Ale Lou Vi
Ile Asp Lys Lys Thr Met Leu Glu Val Glu Arg Ala Leu His Lys Ile Asp Pro
1467 1476 1485 1494 1503
ACT ATC ATT CTC TAC GGC GAA CCG TGG GGT GGS TGG GGT GGS TGG
ACT ATC ATT CTC TAC GGC GAA CCG TGG GGT GGA TGG GGA GCA CCG ATC AGG TTT  The lie lie len Tag Gal and Tag GGA GGA GCA CCG ATC AGG TTT
Thr Ile Ile Leu Tyr Gly Glu Pro Trp Gly Gly Trp Gly Ala Pro Ile Arg Phe
1521 1530 1530
GGA AAG AGC GAT GTC GCC GCC 1539 1548 1557
GGA AAG AGC GAT GTC GCC GGC ACA CAC GTG GCA GCT TTC AAC GAT GAG TTC AGA Gly Lya Ser Asp Val Ala Gly Thr His Val
Gly Lya Ser Asp Val Ala Gly Thr His Val Ala Ala Phe Asn Asp Glu Phe Arg
1575 1584
1575 1584 1593 1602 1611 1620
THE THE MALE CONTRACTOR AND THE PROPERTY OF TH
ASP Ala Ile Arg Gly Ser Val Phe Asn Pro Ser Val Lys Gly Phe Val Het Gly
And Pro Ser Val Lys Gly Phe Val Met Gly

Figure 14C(Continued)

			.mof	oga	mar	iti.	-	Pull	Elar		( 6	0 <b>73</b> )	(c	ont:	lnue	<b>4</b> )	
		162	9		167	0			_								
GG	A TA	C CC	iA A2	NG GA	A AC	CAA	G AT	~ 124 401	,		165	6 .		166	5		1674 TAC
									~ ~G	G GG	r Gr	I GT	r GG	A AC	C AT	AA A	TAC
G1	א זא	æ G1	у Гу	'a G1	u Th	r Ly	s Il	e Ly	s Ar	g G1	y Va	l Val	Gly	 / Se:	r Ile		
		168	3		160	2											
GA	c ca	y yy	Y CI	TK Y	C YY	A AG	T TI	C GC	- - CT	T-GA:	ב ררי דידי	N GAR	C N 1	171	9 • • •		1728 TAC
														AL.	r ati	· AAC	TAC
, A3;	b GT						r Ph	e XI	Le	u Ası	Pro	Glu	Glu	Th		) Ası	Tyr
-		173	7	_ :	174	6		175	5		1764	1		1773	ı		1202
		G 16	r CA	C GA	C AA	CA	C AC	y CIA	TG	GAC	: AAC	אג פֿ	TAC		GCC	GCC	1782 AAA
λl	. A1	a 🗠															
		179											TYI	Lev	Ala	λla	Lys
GCT	' GA'		2 ll:		1800	) 		1809	)		1818	l		1827	,		1836
						1 160	S ACI	C GAL	CAN	GAA	CIG	λλλ	AAC	GCC	CAG	λλλ	1836 CTG
Ala	l Asj	p Lys	Ly:	s Lys	Glu	Tr	Th	Glu	Glu	Glu	Leu	Lys	Asn	λla	Gln	Lys	Leu
		1845	5		1854	ı		1963									
6.2.	. cc.	r GCC	, VII	V CL	. CIC	: λC:	: TC:	געט יו	CCT	Land.	-		~~~	1881			1890 <b>CA</b> G
							·							CAC	AE'S	GGG	CAG
ALA	Gly	/ Ala	Ile	Lev	Leu	Thr	Sex	Gln	Gly	Val	Pro	Phe	Leu	His	Glv	Gly	C1-
		1899														GLy	GIII
CYC	TTC	TGC		100	TAGE			1917			1926			1935		:	1944
		TGC					710	AAC	GAC	AXC	TCC	TAC	AAC	GCC	CCI	ATC	TCG
Asp	Phe	Cys	λrg	Thr	Thr	λεη	Phe	Asn	λsp	Asn	Ser	Tyr	Asn	Ala	 Pro	 Ile	 Ser
		1953			1962			1071									
ATA	AAC	GGC	TTC	GAT	TAC	GAA	AGA	AAA	CTT	CAG	1111.	ATA	CNC	1989	-		1998
														610	110	AAT	TAC
TIE		Gly															
CAC	220	2007 GGT	~~~		2016			2025		2	2034		3	2043		. 2	052
	776	GGT	CIC	ATA	AAA	CIC	λGλ	AAA	CAA	CAC	CCT	GCT	TTC	AGG	CIG	AAA	AAC
His	Lvs	Glv	Len	T)_	Lace	Low											
		Gly 2061															
GCT	GAA	GAG	ATC	AAA	2070	CAC		2079		2	880		2	097		. 2	106
		GAG						GAA	1-1-1.	CTC	ccc	GCC	CCC	λGλ	λGλ	λTA	GTT
Ala	Glu	Glu	Ile	Lys	Lys	His	Leu	Glu	Phe	Leu	Pro	Gly	Gly	Arg	Arg	 Ile	 Val
	- 2	2115		2	2174			1177		_							
GCG	TTC	ATG	CII	AAA	GAC	CAC	CCY	GGT	cci	GAT	CCC	TGG	2 AAA	151 GAC	ÀΤΥ	2 2	160 GTC
Ala	Phe	Met	Leu	Lys	QEA	 His	Ala	Gly	 Gly	 Asp	 Pro	Tro	 Lva	Aen	 Tla	·	
												-			~ * 4		- 47

Figure 14d(Continued)

	There	rotos	70	mar:	ltin	٠ ،	rullı	ılan	450	(60	223)	l e	ont i	<b>5</b> 0-4		
ATT T	2169 AC AAT  Yr Asn 2223	CGA	AAC	2178 TTA	GAG	<b></b>	2187 3 ACA	, 7	TAC	219	5		2205	5	*	2214 TGG
AAT G	C CTT	CIC	AAC	AGC	CAG	አአአ	2241 GCC	GGA	λCA	2250 GAA	בינים	202	2259			2268
GGA AC	ATA A	GAA (	בזכ	CAT	cce	CII	2295 TCC	CCC	TAC	2304 GTT	CTG	TAC	2313	<b></b>		3 ·

Figure 140(Continued)

Figure 15a Thermotoga maritima MSB8 (Clone # 6GP2) Glycosidase

CTT TTA TTG ATC GTT GAG CTC TCT TTC GTT CTC TTT GCA AGT GAC GAG TTC Leu Leu Leu Ile Val Glu Leu Ser Phe Val Leu Phe Ala Ser Asp Glu Phe

GTG AAA GTG GAA AAC GGA AAA TTC GCT CTG AAC GGA AAA GAA TTC AGA TTC Val Lys Val Glu Asn Gly Lys Phe Ala Leu Asn Gly Lys Glu Phe Arg Phe

ATT GGA AGC AAC AAC TAC TAC ATG CAC TAC AAG AGC AAC GGA ATG ATA GAC Ile Gly Ser Asn Asn Tyr Tyr Met His Tyr Lys Ser Asn Gly Met Ile Asp

AGT GTT CTG GAG AGT GCC AGA GAC ATG GGT ATA AAG GTC CTC AGA ATC TGG Ser Val Leu Glu Ser Ala Arg Asp Met Gly Ile Lys Val Leu Arg Ile Trp

GGT TTC CTC GAC GGG GAG AGT TAC TGC AGA GAC AAG AAC ACC TAC ATG CAT Gly Phe Leu Asp Gly Glu Ser Tyr Cys Arg Asp Lys Asn Thr Tyr Met His

CCT GAG CCC GGT GTT TTC GGG GTG CCA GAA GGA ATA TCG AAC GCC CAG AGC Pro Glu Pro Gly Val Pne Gly Val Pro Glu Gly Ile Ser Asn Ala Gln Ser

GGT TTC GAA AGA CTC GAC TAC ACA GTT GCG AAA GCG AAA GAA CTC GGT ATA Gly Phe Glu Arg Leu Asp Tyr Thr Val Ala Lys Ala Lys Glu Leu Gly Ile

AAA CTT GTC ATT GTT GTG AAC AAC TGG GAC GAC TTC GGT GGA ATG AAC Lys Leu Val leu Val Asn Asn Trp Asp Asp Phe Gly Gly Met Asn

CAG TAC GTG AGG TGG TTT GGA GGA ACC CAT CAC GAC GAT TTC TAC AGA GAT Gln Tyr Val Arg Trp Phe Gly Gly Thr His His Asp Asp Phe Tyr Arg Asp

GAG AAG ATC AAA GAA GAG TAC AAA AAG TAC GTC TCC TTT CTC GTA AAC CAT Glu Lys Ile Lys Glu Glu Tyr Lys Lys Tyr Val Ser Phe Leu Val Asn His

GTC AAT ACC TAC ACG GGA GTT CCT TAC AGG GAA GAG CCC ACC ATC ATG GCC Val Asn Thr Tyr Thr Gly Val Pro Tyr Arg Glu Glu Pro Thr Ile Met Ala

TGG GAG CTT GCA AAC GAA CCG CGC TGT GAG ACG GAC AAA TCG GGG AAC ACG Trp Glu Leu Ala Asn Glu Pro Arg Cys Glu Thr Asp Lys Ser Gly Asn Thr

CTC GTT GAG TGG GTG AAG GAG ATG AGC TCC TAC ATA AAG AGT CTG GAT CCC Leu Val Glu Trp Val Lys Glu Met Ser Ser Tyr Ile Lys Ser Leu Asp Pro

AAC CAC CTC GTG GCT GTG GGG GAC GAA GGA TTC TTC AGC AAC TAC GAA GGA Asn His Leu Val Ala Val Gly Asp Glu Gly Phe Phe Ser Asn Tyr Glu Gly

TTC AAA CCT TAC GGT GGA GAA GCC GAG TGG GCC TAC AAC GGC TGG TCC GGT Phe Lys Pro Tyr Gly Glu Ala Glu Trp Ala Tyr Asn Gly Trp Ser Gly

GTT GAC TGG AAG AAG CTC CTT TCG ATA GAG ACG GTG GAC TTC GGC ACG TTC Val Asp Trp Lys Lys Leu Ser Ile Glu Thr Val Asp Phe Gly Thr Phe

CAC CTC TAT CCG TCC CAC TGG GGT GTC AGT CCA GAG AAC TAT GCC CAG TGG His Leu Tyr Pro Ser His Trp Gly Val Ser Pro Glu Asn Tyr Ala Gln Trp

GGA GCG AAG TGG ATA GAA GAC CAC ATA AAG ATC GCA AAA GAG ATC GGA AAA Gly Ala Lys Trp Ile Glu Asp His Ile Lys Ile Ala Lys Glu Ile Gly Lys

CCC GTT GTT CTG GAA GAA TAT GGA ATT CCA AAG AGT GCG CCA GTT AAC AGA Pro Val Val Leu Glu Glu Tyr Gly Ile Pro Lys Ser Ala Pro Val Asn Arg

ACG GCC ATC TAC AGA CTC TGG AAC GAT CTG GTC TAC GAT CTC GGT GGA GAT Thr Ala lle Tyr Arg Leu Trp Asn Asp Leu Val Tyr Asp Leu Gly Gly Asp

GGA GCG ATG TTC TGG ATG CTC GCG GGA ATC GGG GAA GGT TCG GAC AGA GAC Gly Ala Met Phe Trp Met Leu Ala Gly Ile Gly Glu Gly Ser Asp Arg Asp

GAG AGA GGG TAC TAT CCG GAC TAC GAC GGT TTC AGA ATA GTG AAC GAC GAC Glu Arg Gly Tyr Tyr Pro Asp Tyr Asp Gly Phe Arg Ile Val Asn Asp Asp

AGT CCA GAA GCG GAA CTG ATA AGA GAA TAC GCG AAG CTG TTC AAC ACA GGT Ser Pro Glu Ala Glu Leu Ile Arg Glu Tyr Ala Lys Leu Phe Asn Thr Gly

GAA GAC ATA AGA GAA GAC ACC TGC TCT TTC ATC CTT CCA AAA GAC GGC ATG Glu Asp Ile Arg Glu Asp Thr Cys Ser Phe Ile Leu Pro Lys Asp Gly Met

GAG ATC AAA AAG ACC GTG GAA GTG AGG GCT GGT GTT TTC GAC TAC AGC AAC

Figure 15b (continued)

Glu Ile Lys Lys Thr Val Glu Val Arg Ala Gly Val Phe Asp Tyr Ser Asn

ACG TTT GAA AAG TTG TCT GTC AAA GTC GAA GAT CTG GTT TTT GAA AAT GAG Thr Phe Glu Lys Leu Ser Val Lys Val Glu Asp Leu Val Phe Glu Asn Glu

ATA GAG CAT CTC GGA TAC GGA ATT TAC GGC TTT GAT CTC GAC ACA ACC CGG Ile Glu His Leu Gly Tyr Gly Ile Tyr Gly Phe Asp Leu Asp Thr Thr Arg

ATC CCG GAT GGA GAA CAT GAA ATG TTC CTT GAA GGC CAC TTT CAG GGA AAA Ile Pro Asp Gly Glu His Glu Met Phe Leu Glu Gly His Phe Gln Gly Lys

ACG GTG AAA GAC TCT ATC AAA GCG AAA GTG GTG AAC GAA GCA CGG TAC GTG Thr Val Lys Asp Ser Ile Lys Ala Lys Val Val Asn Glu Ala Arg Tyr Val

CTC GCA GAG GAA GTT GAT TTT TCC TCT CCA GAA GAG GTG AAA AAC TGG TGG Leu Ala Glu Glu Val Asp Phe Ser Ser Pro Glu Glu Val Lys Asn Trp Trp

AAC AGC GGA ACC TGG CAG GCA GAG TTC GGG TCA CCT GAC ATT GAA TGG AAC Asn Ser Gly Thr Trp Gln Ala Glu Phe Gly Ser Pro Asp Ile Glu Trp Asn

GGT GAG GTG GGA AAT GGA GCA CTG CAG CTG AAC GTG AAA CTG CCC GGA AAG Gly Glu Val Gly Asn Gly Ala Leu Gln Leu Asn Val Lys Leu Pro Gly Lys

AGC GAC TGG GAA GAA GTG AGA GTA GCA AGG AAG TTC GAA AGA CTC TCA GAA Ser Asp Trp Glu Glu Val Arg Val Ala Arg Lys Phe Glu Arg Leu Ser Glu

TGT GAG ATC CTC GAG TAC GAC ATC TAC ATT CCA AAC GTC GAG GGA CTC AAG Cys Glu Ile Leu Glu Tyr Asp Ile Tyr Ile Pro Asn Val Glu Gly Leu Lys

GGA AGG TTG AGG CCG TAC GCG GTT CTG AAC CCC GGC TGG GTG AAG ATA GGC Gly Arg Leu Arg Pro Tyr Ala Val Leu Asn Pro Gly Trp Val Lys Ile Gly

CTC GAC ATG AAC AAC GCG AAC GTG GAA AGT GCG GAG ATC ATC ACT TTC GGC Leu Asp Met Asn Asn Ala Asn Val Glu Ser Ala Glu Ile Ile Thr Phe Gly

GGA AAA GAG TAC AGA AGA TTC CAT GTA AGA ATT GAG TTC GAC AGA ACA GCG Gly Lys Glu Tyr Arg Arg Phe His Val Arg Ile Glu Phe Asp Arg Thr Ala

Figure 15C(continued)

GGG GTG AAA GAA CTT CAC ATA GGA GTT GTC GGT GAT CAT CTG AGG TAC GAT Gly Val Lys Glu Leu His Ile Gly Val Val Gly Asp His Leu Arg Tyr Asp

GGA CCG ATT TTC ATC GAT AAT GTG AGA CTT TAT AAA AGA ACA GGA GGT ATG Gly Pro Ile Phe Ile Asp Asn Val Arg Leu Tyr Lys Arg Thr Gly Gly Met

TGA 1991

END

Figure 15d(continued)

# Figure No. 16/Thermotoga maritima MSB8(6gb4)

_			- A	- AI	. C GA	C CT	G AA	r GG:	TT(	C TG	3 AG	C GT	r ago	GA	T AA	C GA	A GG	G 10	n		
1	Me	t Ly	's Ai	rg Il	e As	p Le	u Ası	Gly	y Phe	e Tra	Se	. Va	l Arc	n Ae	- No	- Cl		G AG	A 11	TT	CG (
										_				, ,,,,	PAS	11 GT	u Gi	y Ar	g Ph	e s	er :
61	TT	T GA	A GO	3G AC	ערט ער	·C -C-C-															
21	Ph	e G1	(1	G AC	- GI	6 CC	A GGC	GT	GTO	CAC	G CZ	A GA	CIC	GT	C AG	A AA	A GG	т ст	T CT	тс	CA 12
		c GI	u G1	y Th	r va	1 Pro	G13	' Val	Val	Glr	Ala	Ası	Let	va:	LAr	g Ly:	s G1	y Le	u Le	u P	ro 4
121	CA	c cc	G TA	C GT	T GG	G AT	AAC	GAA	GAT	СТС	<b>T</b> TC		. car								
41	Hi.	s Pr	o Ty	r Va	1 G1	y Met	Asn	Glu	Acr	1.01	Dha	· And	GAA	AT	A GA	A GA	: AG	A GA	GTG	G A	TC . 18
						•				Deu	Pile	. TAS	GIU	llle	: G1:	ı Ası	) Ar	3 G1	ı Tr	p I	le 6
181	Th/	c ca	c																		
61	~	- 67	G AG	G GA	G TT	C GAG	TTC	AAA	GAA	GAT	GTG	AAA	GAG	GGG	GAJ	CG1	GTO	GA:	ירם יו	_ G1	TT 24
61	1y:	r GI	u Ar	g Gl	u Ph	e Glu	Phe	Lys	Glu	Asp	Val	Lys	Glu	Gly	Glu	Arc	Va]	Ast	1.61	. V.	1 0
														_		•		,	- 261	4 V C	11 8
241	TT	r ga	G GG	C GT	C GA	C ACG	CTG	TCG	GAT	CTTT	ጥልጥ	CTC									
81	Phe	e Gl	u Gl	y Vai	l Ası	o Thr	Len	Sar	200	1/2.3	7		AAL	GGI	GTT	TAC	CTI	: GG	AG(	) AC	C 30
				y Vai				561	vsh	vai	lyr	Leu	Asn	Gly	Val	Tyr	Let	Gly	' Sea	Th:	r 10
301	CAZ			1							•										
101	GA,	GA	CAT	G TT	CATO	GAG	TAT	CGC	TTC	GAT	GTC	ACG	AAC	GTG	TTG	AAA	GAA	AAC	AAT	CA	C 35
101	GIL	ı Ası	o Me	t Phe	e Ile	e Glu	Tyr	Arg	Phe	Asp	Val	Thr	Asn	Val	Leu	Lys	Glu	Lvs	AST	. Wi	s 120
																•		-,-		• •••	
361	CTG	AAC	GT	J TAC	TATA	AAA	TCT	ccc	ATC	AGA	התים	CCG		» ~~	~~~						•
121	Leu	Lys	Va!	1 туг	· Ile	Lvs	Ser	Pro	Tle	Ara	Val	D=-	7	ACI	Cre	GAG	CAG	AAC	TAC	GG	G 420
				-	_	-,-	-,		116	Ar 9	val	PIO	гÀг	Thr	Leu	Glu	Gln	Asn	Tyr	Gl	y 140
421	GTC																				•
141	Val		- 666	GGI	CCI	GAA	GAT	ECC	ATC	AGA	GGA	TAC	ATA	AGA	AAA	GCC	CAG	TAT	TCG	TA	C 480
	461	Ten	r GT	/ Gly	Pro	Glu	Asp	Pro	Ile	Arg	Gly	Tyr	Ile	Arg	Lys	Ala	Gln	Tyr	Ser	Tv	r 160
																		•		-7	
481	GGA	TGG	GAC	TGG	GGT	GCC	AGA	ATC	GTT	ACA	AGC	GGT	) TTT	TCC							
161	Gly	Trp	Asp	Trp	Gly	Ala	Arq	Ile	Val	Thr	Ser	Gly	710	7-	****	-	GTC	TAC	CTC	GA	5 540
							•					Oly	116	11D	гÀ2	PTO	Val	Tyr	Leu	Gl	180 נ
541	GTG	TAC	ACC.																		
	Val	Tier	7	GCA	CGT	CTT	CAG	GAT	TCA	ACG	GCT	TAT	CTG	TTG	GAA	CTT	GAG	GGG	AAA	GA:	600
		TYL	Arg	Ala	Arg	Leu	Gln	Asp	Ser	Thr	Ala	Tyr	Leu	Leu	Glu	Leu	Glu	Gly	Lys	Ast	200
601	GCC	CII	GTG	AGG	GTG	AAC	GGT	TTC	GTA	CAC	GGG	GAA	GGA	<b>ד</b> ממ	CTC	. ~~					
201	Ala	Leu	Val	Arg	Val	Asn	Glv	Phe	Val	His	Gly	G3	C1	<b>N</b>	CIC	ATT	GTG	GAA	GTT	TAT	660
							•				O.y	GIU	GIY	ASN	reu	Ile	Val	Glu	Val	Tyr	220
661	GTA	220	CCT	~~~		*															
221	Val	300	991	GAA	AAG	ATA	GGG	GAG	TTT	CCT	GTT	CTT	GAA	AAG	AAC	GGA	GAA	AAG	CTC	TTC	720
	· aı	ASN	GIY	Glu	Lys	Ile	Gly	Glu .	Phe	Pro '	Val	Leu	Glu	Lys	Asn	Gly	Glu	Lvs	Leu	Phe	240
					•			•													
721	GAT	GGA	GTG	TTC	CAC	CTG	AAA (	GAT (	GTG	י בגב	י בידי	TCC ·	ጥስጥ		<b>T</b> C •						
241	Asp	Gly	Val	Phe	His	Léu	Lvs	Asp 1	Val 1	live 1		- <del></del>	-A1		TGG -	AAC	GTG	GGG	AAA	CCG	780
					_		-10	٠٢	· u.z	-ys 1	oeu '	TID.	ıyr	Pro	Trp	Asn	Val	Gly	Lys	Pro	260

•	781	TAC	CTG	TAC	GAT	TTC	مست															•
• :	261 :	Tyr	Leu	Tvr	Acn	Dha	GII	TTC	GTG	TTG	AAA	GAC	TTA	AAC	GGA	GAG	ATC	TAC	DC N	- C N 7	GAA	
				- 7 -	nsp	File	vai	Phe	Val	Leu	Lys	Asp	Leu	Asn	Gly	Glu	Ile	Tvr	7-0	CAL	GAA	840
															•			- , -	λιg	GIU	Glu	280
	341 /	VAG 2	AAA	ATC	GGT	TIG	AGA	AGA	GTC	AGA	ATC	سلن	CAC	CNO								
-	81 I	ys 1	Lys	Ile	Gly	Leu	Arg	Arg	Val	AGA Arg	Tla	1/21	CAG	GAG	CCC	GAT	GAA	GAA	GGA	AAA	ACT	900
								_			116	VAI	GIN	Glu	Pro	Asp	Glu	Glu	Gly	Lys	Thr	300
9	01 T	TC A	ATA '	TTC	GD N	200																
3	01 P	he I	le i	Phe i	Cl.	71.	AAC	GGT	GAG	AAA Lys	GTC	TTC	GCT	AAG	GGT	GCT	AAC	TCC	7 mm			
		_			GIU	TTE.	Asn	Gly	Glu	Lys	Val	Phe	Ala	Lys	Glv	Ala	Acn	7	wii.	CCC	TCA	960
														•	3		A311	irb	Tie	Pro	Ser	320
	61 G	AA A	AC A	ATC (	CTC :	ACG '	TGG ·	TTG	AAG	GAG Glu	GAA	Chm.										
3:	21 G	lu A	sn 1	le I	Leu :	Thr :	Trp :	Leu	Lve	Glu	C1	GAI	TAC	GAA .	AAG	CTC (	STC .	AAA	ATG	GCA	AGG	1020
					•		•		-, -	Glu	GIU .	Asp	Tyr (	Glu	Lys :	Leu 1	Val :	Lys :	Met	Ala	Ara	340
102	21 AC	T G	מר ש	ልጥ ክ	· TC •	·																340
34	1 Se	T A	ים בי אים בי	W		VAC A	ATG (	CTC	AGG	GTC :	TGG (	GGA (	GA (	GA J	ATC 1	ראר נ	י ממי					
		- 1.	-a A	sn M	let A	sn M	let I	leu )	yrg .	Val 1	Trp (	Gly (	ily (	ilv i	120	····		AGA (	∂AG .	ATC	TTC	1080
																						360
108	1 TA	C AC	SA C	TC T	GT G	AT G	AA C	TC o	GT :	ATC A		-										
36	1 Ty	r Ar	g L	eu C	ys A	sp G	lu ī.	en c	27.0			1G 1	GG C	'AG G	AT T	TC A	TG T	AC G	CG 1	GT (	CTT	1140
					-	•			ту.	ile M	iet V	al T	zb G	ln A	sp P	he M	et T	yr A	la c	Cvs I	Leu	380
114																						200
38:	1 61	1 72	·- ~	- 6	AT C	AT C	TT C	CG T	GG I	TC A	GA A	AA C	TC G	CG A	AC G	A A C	, .	<b>~</b> .				
	- 01	- 1y	I PI	O As	sp H	is L	eu P	ro T	rp P	he A	rg L	ys L	eu A	la A	SD G	1	MG (G	CA A	GA A	AG A	TT	1200
																						400
1201	L GTO	AG:	AA A	A CI	C AC	A T	C C	AT C	CC T	CC A:	T-T				•		-					
401	. Val	Arg	g Ly	s Le	u Ar	g 1\	r Hi	S Pi	rn 8	er II	1. 6.	11 C.	C TO	G T	GC GC	EA A	ريز ع <u>ا</u>	.C G	A A	AC A	AC	1260
										1.	re va	al Le	u Tr	љ с ^у	's G]	y As	n As	n G	u A	sn A	sn	420
1261	TGG	GGZ	4 T-T	c çv	T 0.																	
421	Tro	Glv	, Dh	- GA	- GA	A TG	G GG	A A	AT A	rg gc	C AC	a aj	A GT	G GA	T GG	T AT	ממ "	C C				
	•	1		- AS	D GI	u Tr	p Gl	y As	n Me	et Al	a Ar	g Ly	s Va	l As	p G1	v 71	- 10	- T		A A	AC.	1320
1271																,	c ns	n Te	u Gl	у Аз	SI)	440 .
1321	AGG	CIC	TAC	CT	C TT	C GA	T TT	T CC	T G#	G AT	יד דים	T 60	c		_							
441	Arg	Leu	Tyz	Lei	u Ph	e As	p Ph	e Pr	o G1	u 71	- ··	- 33	C GA	A GA	A GA	c cc	G TC	C AC	T CC	C TA	T	1380
										u Il	c cy	2 AT	a GI	n Cl	u As	p Pr	o Se	r Th	r Pr	o Ty	'r	460
1381																						
461	Trp	Pro	Ser	Ser		· IA	- GG(	. GG	r ga	A AA	A GC	G AA	AG(	GAJ	A AA	G GAJ	A GG	4 C2	~ 20	~ ~~		
	_			Jei	PIC	y Tyr	. G1 ²	/ Gl	/ G1	u Ly:	s Ala	a Ası	Sea	Gli	Lve	G)	. 61.		- 70	G CA		1440
1441	-	_				٠									•	-	- 01)	, AS	Ar	д ні	S	480
1441	GTC	TGG	TAC	GTG	TGG	AGI	GGC	TGO	ATO	G AAC : Asn	TAC	- CA 2			_							
481	Val	Trp	Tyr	Val	Trp	Ser	Gly	Tre	Mei	- Ac-		- 074	AAC	TAC	GAA	AAA A	GAC	ACC	GG	A AG	3 1	500
							•	- 4	• • • • • •	- ASI	ııyı	GIU	Asn	Tyr	Glu	Lys	Asp	Thr	G1)	/ Arg	3	500
1501	TTC .	ATC	AGC	GAO	مستحد														-	•		
501	Phe	Ile	Ser	G)	TIT.	GGA	TTT	CAG	GGT	GCT	. ccc	CAT	CCA	GAG	ACG	ATA	GAC	حسم				
	Phe :			210	LUG	Gly	Phe	Gln	Gly	Ala	Pro	His	Pro	Glu	Thr	T1.	Class	110	111	TCA	. 1	560
1561															- •••	T16	GIU	rhe	Phe	Ser		520
	AAA (	CC (	GAG	GAA	AGA	GAG	ATA	TTC	САТ	ככר	CTC	h ===										
521	AAA (	, ro	Glu	Glu	Arg	Glu	Ile	Phe	Hie	Dro	V-3	AIG	CIG	AAG	CAC	AAC	AAA	CAG	GTG	GAA	. 10	520
	Lys I												Deu	Lys	His	Asn	Lys	Gln	Val	Glu	9	540
								F	igu	re l	6 <b>b</b> (d	ont	inue	d)								

1623	GG	A CA	G GA	A AG	A TT	~ ~ ~			_												
541	. G1	י מו	n (1)			G AT	CAG	3 TTC	ATA	TTC	GGA	AA:	TI.	r GG)	AA F	G TG	T AA	A GA	т тт	C GAC	
		, GI.	ıı Gı	u Ar	g Le	u Ile	e Arc	Phe	: Ile	Phe	Gly	Ası	n Phe	e G1v	Lv	5 Oz				C GAC	
															-,	J Cy	s Ly	S AS	p Pn	a Asp	56
1681	AG:	r.TT	T GŢ	G TA	r cr	G TC	CAC	: CTC	. איני												
561	Se	Phe	Va.	l Tv:	t Lei	Ser	. 61-			- CAG	GCG	GAG	GCC	ATC	: AAC	TT	GG:	r GT	r gaj	A CAC	1740
				•			. 611	reu	Asn	Gln	Ala	Glu	Ala	Ile	Lys	Phe	: Gly	Va:	Gli	A CAC	580
1741																					380
	-	. CG/	AG	2 AGC	AAC	TAC	AAA	ACG	GCC	GGC	GCT	CTC	TTC	TGG	CAC	· mmc				TGG	
581	Trp	Arg	Ser	Arg	Lys	Tyr	Lys	Thr	Ala	Gly	Ala	Leu	Dhe	T	CAG		AAC	GAC	AGC	Trp	1800
										•			FILE	irp	GID	Phe	Asn	Asp	Ser	Trp	600
1801	CCG	GTC	TTC	AGC	TGG	TOO	CCN													•	
601	CCG Pro	Val	Phe	Ser	T~	. Ci-	GCA	GTC	GAT	TAC	TTC	AAA	AGG	CCC	AAA	GCT	CTC	TAC	TAC	TAT	1860
	Pro				110	SEL	ALA	Val	Asp	Tyr	Phe	Lys	Arg	Pro	Lys	Ala	Leu	Tvr	Tyr	Ture	
1861																					620
	GCG	AGA	AGA	TTC	TTC	GCT	GÁA Glu	GTT	CTA	CCC	GTT	TTG	DAG	220	202				•		
621	Ala	Arg	Arg	Phe	Phe	Ala	Glu	Val	Leu	Pro	Va 1	Leu	7	7	AGA	GAC	AAC	AAA	ATA	GAA	1920
											-	Deu	Буз	rys	Arg	Asp	Asn	Lys	Ile	Glu	640
921	CTG	CTG	GTG	GGT	CAC	000															
641	Leu	Len	77.1	23	ano.	- CGA	TCT	GAG	GGA	GAC .	AAA	AGA	AGT	CTC	TCT	CAG	GCT	TGC	AGC	СТА	1980
			147	GIY	Glu	Arg	Ser	Glu	Gly	Asp :	Lys .	Arg	Ser	Leu	Ser	Gln	Ala	Cvs	Ser	Tan	
																					660
981	CGA Arg	GAA	gaa	GGG	AGA	AAA	GGT	ATT	CGA :	. מממ	280 .	T	~>~								
661	Arg	Glu	Glu	Gly	Arg	Lys	Glv	Tle	ATO 1			TIM	CAG	AAC	GGT	ACT	CCC	AGC	AGA	CGG	2040
					_	•	,		y	uys A	sp.	ren .	Gln .	Asn	Gly	Thr	Pro	Ser	Arg	Arg	680
041	TGT																				
						20	55														٠.
	Cys	GT (I	Pne.	Gly	End	68	5														

Figure 16 C(continued)

# Figure No. 17 Bankia gouldi (37gp4)

	1 A	T																							
	- A	IG A	AA J	\AA	AAT	CIX	CI	A AT	G I	TT	AAA	AG	CT	T AC	G T	AT (	СТА	CC-T						CTG	
:	1 M	et L	ys I	ys	Asn	Let	1 Le	u Me	t P	he	Lvs	Arc	Lei	ı Th	~ ~	1		-	- 110	11	r TT	A A	rg	CTG Leu	60
													,			7 = 1	Leu	Pro	Leu	Ph	e Le	u Me	ŧ	Leu	20
6:	~ ،	ייי ייי	Ch C	-																				•	
21	- <b>-</b> .	TC T		.1A	AGT	TC	GT	A GC	TC	AA	TCT	CCJ	GT	A GA	A AZ	A C	CAT	GGC	CGT	TT	. (2)	A C7	~	~~~	
2.	L 1.0	eu S	er L	eu	Ser	Ser	Va	l Al	a G	ln .	Ser	Pro	Va:	Gl	u L	rs H	i s	Gly	7.0	7.00				GAC	120
															- 4			<b>U</b>	ÆΥ	ret	I GII	o Va	I	Asp	40
121	G	SA AS	AC C	GC .	ATT	CTT	יממ י	ר כר	ci m	~~~ <i>,</i>	~~ "	~~~							•						
41	GC G1	y A	n A	ro	Tle	T.011	N			-1 (		GAA	ATT	· AC	3 AG	CI	ATT	GCT	GGT	AAC	AG	CT	C	TTT	180
		y A		-3			, vai	1 MI	۵.5	er	31y	Glu	Ile	Th	Se	r L	.eu	Ala	Gly	Asn	Ser	Le	u	Phe	60
181	TG	G AC	ST A	AT (	GCT	GGA	GAC	AC	C TO	c c	SAT	TTT	TAI	' AA1	r GC	A G	. A A	л <del></del>	Cmm	~					
61	Tr	p Se	r A	sn i	Ala	Gly	aak	Th	r Se	r )	Asp	Phe	Tvr	Agr	. 21	·· -		W	G11	GAT	TTT	TT.	A	GCA	240
											•		- , -			a	TU	Inr	Val	Asp	Phe	Le	u .	Ala	80
241	GA	A A	C T		יד ת ב	3.00				_															
81	Gl	A AA u As	. T		~	AGC	TCA	CT.	r Al	T	\GA	ATA	GCT	ATG	GG	C G	TA	AAA	GAA	AAT	TGG	GA:	r	GGC ·	300
	-	u As	1.	rb ,	ısn	Ser	Ser	Lei	11	e A	rg	Ile	Ala	Met	G1	y V	al	Lys	Glu	Asn	Tro	Acr	- ·	23.4	
																									100
301	GG.	A AA Y As	T G	3C 1	TAT	ATT	GAT	AGT		G C	AG	GAG	C22	CD 2	~~										
201	G1	y As	n G	ly 1	yr	Ile	Asp	Ser	. Pr	n G	ים ו	-21.0 G1.0	C1-	02		ىخدا	AA.	ATT	AGA	AAA	GTT	ATT	C	GAT	360
							. •				~	<b>01</b> 0	GIII	GIU	Ala	L	ys :	Ile	Arg	Lys	Val	Ile	}	\sp	120
361	GC	A CC	T 3 4	~ _																			•	•	
121	ינמ	A GC	:	T G	CT.	AAC	GGC	ATA	TA	TG	TA .	ATA	ATA	GAC	TGC	; C	AC I	ACT	CAC	GAA	GCA	GAG	: "T	מידי	420
	NT.	a Al	a 11	.e A	la.	Asn	Gly	Ile	Ty	r V	al :	Ile	Ile	Asp	Tri	Hi	is 1	Thr :	His	G)	Ala	61			
																									140
422	. TAC	AC	A GA	T G	AG (	GCT	GTT	GAC	T-T-	T 3"	, 1		202												• •
141	Tyr	Th	r As	рG	lu į	Ala	Val	Asn	Ph	. D		76	NUM	AIG	GCA	. GA	rc c	TA	TAC	GGA	GAT	ACT	C	CC .	480
141								-100	F 1.10	. P.		mr	Arg	Met	Ala	As	p I	eu '	Tyr	Gly	Asp	Thr	P	ro	160
481																									
161	No.	GT/	AT	G T	AT (	SAA	ATT	TAT	AA	G	AG C	CT.	ATA	TAC	CAA	AG	TI	GG (	CT (	STT	ATT	DAG	2	h T	540
*01	ASI	Va]	. Me	t T	Yr (	Glu	Ile	Tyr	Ası	G	lu F	,ro	Ile	Tyr	Gln	Se	r T	י סבי	200 1	7=1	77.0	****		 w1	
																		-p .		- 41	116	гÀè	A.	БŊ	180
541	TAT	GC3	GA	G C	AA C	STA .	ATT	GCT	GC1	י א	רא כ	·												•	
181	Tyr	GCA Ala	Gl	u G	ln v	/a 1	Tle	717	001			.GI	rer	AAA	GAC	CC	A G	AT ;	LAT :	TA	ATA	ATT	G'	TA	600
		Ala						710	Gly	11	.е д	rg :	ser	Lys	Asp	Pr	O A	sp A	sn 1	eu	Ile	Ile	V	al	200
601																									
201	GGI	ACT Thr	AG	CA	AT T	AT :	CT	CAG	CAA	GI	T G	AT (	STA :	GCA	TCA	GC	A G	AC C	י בי	י מידי					
201	Gly	Thr	Sea	r As	T C	yr :	Ser	Gln	Gln	Va	1 A	sp t	/al .	Ala	Ser	A1:	a 1	co 0			101	GAT	AC	-T	660
																7121	<b>u</b> A	sp F	TO 1	le :	ser.	Asp	Tì	ır	220
661	AAT	GTG	GCZ	. TA	4 TJ	CT 7	מיזיין	C 11 T	<del></del>		<b>.</b> -														
221	AAT Asn	Val	Ala	ינים	· ·	h		us -	LIT	TA	T G	CA (	CA '	ITT .	AAC	CCC	G C	AT G	AT A	AC :	TTA	AGA	AA	\T	720
	Asn	-		- <b>-</b> y	- 1	•++ I	eu i	n1S	Phe	Ту	T A	la 🏞	la	Phe .	Asn	Pro	> H:	is A	sp A	sn 1	Leu i	Ara	As	sn.	240
									•																
721	GTA Val	GCA	CAG	AC	A G	CA 1	TA	GAT	AAT	AA	T G	rr c	CT -	، تاليا	مدسوسل	~~~	n .	~ .							
241	Val	Ala	Gln	Th	r A	la L	eu 1	Asp	Asn	Ası	n V:	C	1.		LIT	G I I	. A(	≟A G	AA T	GG (	GT /	ACA	AT	T	780
								-				4	4 d 1	eu :	rne	va1	L T	ar G	lu T	TD C	เวา	rh-	τī	_	260

	_																						. 4	
		81	TTA	AAT	ACC	GGA	CAA	GGA	GAA	CCA	CAC						-	٠.						
	2	61	Leu	Asn	Thr	Glv	C1 m	G3			GAC	AAA	GAA	AGC	ACT	AAT	C AC1	TGG	ATO	GC	СТ	TT TI he Le	'C 040	
						,	GIII.	GIY	GIU	Pro	Asp	Lys	Glu	Ser	Thr	Asn	Thr	· T	Mar				G 840	
																		***	Met	. Als	2 P.	ne Le	u 280	
	84	11	AAA	GAA	AAA	CCT	ስ ሞ አ																•	
	28	31	T.tre	C1	•		nin ,	WG I	CAC	GCT	AAT	TGG	TCT	TTG	AGT	GAC	444	CCT	. Andread	٠		AA AC		
		-	<b>-</b> ,	<b>G1</b> II	ràs	GLY	Ile :	Ser	His	Ala	Asn	Tro	Ser	Len	Sa=				111	CCI	G	AA AC	A 900	
												•			Set	Asp	Lys	Ala	Phe	Pro	G3	lu Th	300	
	90																							
•	20				GTA	GIT (	CAA (	GCA (	GGA	CAA	GGT	GTA	TCT	GGT	477	y and						A GC		
	30	1 (	ily :	Ser '	Val '	Val (	3ln ;	Ala (	Gly i	Gln	ดาจ	Va 1	Co	<b>a</b> 2.		WII	AGC	AAT	AAA	CTT	AC	A GC	960	
									•		,	•44	SEL	GIÀ	Leu	Ile	Ser	Asn	Lys	Leu	Th	A GC	320	
	96																							
		_ '	CT. (	GT (	GAA A	TT C	TA A	LAA A	VAC I	ATC :	ATC (	CAA	AAC .	TCC (	~ > ~							A CCI		
	32	1 5	Ger (	ly (	3lu 1	le v	al L	vs 1	sn .	r1.	71.	~``	-	166 (	SAT.	ACA	GAG	ACC	TCT	ACA	GG	A CCI	1020	
											TTE (	oin .	ASD :	Trp 1	Asp '	Thr	Glu	Thr	Ser	Thr	G1	A CCI y Pro	340	
	100																							
	102	ı A	AA A	CA A	CA C	AA I	GT A	GT A	CT A	TA C	:AA 7	ו יייניים	, 444									A GCA		
	34:	l L	ys .T	hr 1	hr G	ln c	vs s	e :- T	h- +				-111	AGA C	SCT (	<b>GCA</b>	ATG	GAA	ACA	GCA	CA	A GCA	1080	
							,	1	117 1	Te G	iiu (	ys I	[le ]	ra y	la 2	Ala .	Met	Glu	Thr	Ala	G) •	A GCA	2000	
																							360	
	1081	G	GA G	AT G	AA A	TT A	TA A	TT G	ככ כ	CT C	ת מכו	20 5										GCC		
	361	G	ly A	sp G	lu I	le T	) e T'	א מו			<b>.</b>	MC I	AC A	AT T	TT C	CAA (	GAC .	AAG .	ATA	CAA	GG1	GCC Ala	1140	
				-	_		1.	TE M	ıa P	ro G	ly A	sn I	уг А	sn P	he G	iln i	Asp :	Lys	Ile i	G) n	G) v			
																							380	
	1141	T	T A	AC C	ST A	ST G	TT TA	C C	יי יי	אד כ	CT N	~~ ~										ATA		
	381	Pì	e A	נא מפ	ra se	er Va	al 1%	·			GI A	GT G	CTA	AC G	ga a	AC A	AGT J	ACA A	AAC (	CCT ;	ATT	ATA	1200	
						"	al Ty	YE LIE	u T	yr G	ly s	er A	la A	sn G	ly A	sn s	Ser 1	hr 2	Asn 1	Dro '	71.	73-		
																							400	
	1201	TI	A AC	A GC	SC GZ	A AC	C GC	יים ידי	יה הי		-													
	401	Le	u Ar	9 G1	ע פו	,, 60		\	~ ~	ייר כי	-1 (	TT G	IT T	IC IC	CA GO	GA I	TA G	AT I	AT A	VAC 2	AAT	GGC	1260	
					, 01	u 5e	r Al	a Th	I AE	n Pr	O Pr	o Va	al Pi	ne Se	r G	ly L	eu A	Sp T	מ יינו			23		
																							420	
	1261	TA	C CI	A TI	A AG	T AT	T GA. e Gl:	A CC	מים יוד	~ ~ ~														
	421	Ty	r Le	u Le	1) Se	- T)	- 01.		- GA	1 14	T TG	G A	TA TA	TAA	A GA	A TA	TA G	AG T	TT A	AA A	CT	GGG	1320	
		-				- 11	e Gl	n GT	y As	р Ту	T Tr	P As	n Il	e Ly	S As	p I	le G	ם וו	he t		·			
																							440	
	1321	TC	C AA	A GG	T AT	r Gt	r Cr	T GN														•		
	441	Sea	Lv	s Gl	. T1			· GA	- AA	r rc	T AA	T GG	T AG	AA T	A TT	A A	AA AA	AC C	TT G	TT G	TT	ראר	1380	
			_,	- 01	, 11,	= va.	l Let	ı Ası	As:	n Se	r As:	n Gl	y Se	r Ly:	s Le	u L	/s A	en Ta	37	- 1				
																							460	
1	381	GAT	ATT	GGJ	A GAZ	CAR	GCI				_													
	461	Ast	Ile	. G1,	. 61.			AL I	CAC	TT	G CG	r GA	T GG	A TC	r ag	C AA	T AZ	TAC	T AT	רא כו	ስ ጥ <i>የ</i>	مهتات	3440	
		•		. 01)	GIU	GIU	Ala	Ile	His	Let	Arg	a As	9 Gl	y Ser	Se	r Ac	n be	n 6-	- *1		••		1440	
									•				•			- 7.0	M AS	Se	E 11	le As	sp (	Sly	480	
1	441	TGC	ACT	ATA	TAC	" אא	202																	
	481	Суя	Th-	71~	~~~~		ACA Thr	UG T	AGA	ACT	AA	CCI	r GG1	TTI	GG	r ga	A GG	т тт	מד ב	י. דים ידי	· A ·	200	3500	
		• -		**6	TAL	ASD	Thr	Gly	Arg	Thr	Lys	Pro	Gly	, Phe	Glv	/ G1	ום וו	- • - • •					1500	
																							500	
1	501	TCA	GAT	AAA	GCA	( D D	~~~	<b></b>																
!	501	Ser	Asn	Tare			CAT His	GAC	ACT	TAT	GAA	AGA	GCT	TGT	AAC	: AA	TAA	ר ארי	ጥ አጥ	T 0.				
			p	Lys	GIA	Gln	His	Asp	Thr	Tyr	Glu	Arg	Ala	Cvs	Asn	. A = -	n »		- AI	. GA	л А	ML	1560	
												-		- , -		- 451	· AS	Th:	r Il	e Gl	u A	sn	520	
15																								
5	21 (	Cys	Th-	Ve 1		-	AAT Asn	GTA	ACA	GCA	GAA	GGC	GTA	GAT	GTT	' AAC	GA:	, GO	r >~	,				
			4	AGT	GIA	Pro	Asn	Val	Thr	Ala	Glu	Glv	Va)	Asn	Val	7		- 56,	AC	n AT	Äن	AC	1620	
							Asn					- 3		55	441	υyε	• GIL	Gly	Th	r Mei	t A	sn	540	

Figure 17b(continued)

16	21 A	CT ;	ATT	ATZ	A AG	A AZ	AT TO	יר כיי	·~				•												
S	21 A 41 T	hr :	Ile	Ile	. Ar	g As	in Cv	e va	3 27	T TC	T GC	'A GA	VA G	GA A	TT	rca	GGA	GAA	AA	T A	GC :	rca	GAT	168	^
	41 T					•	<b>-</b> ,	J va	- PN	e Se	r Al	a Gl	u G	ly I	le s	Ser	Gly	Glu	As	n Se	er s	er	Asp	560	
16																									,
	61 A	1	14.	WI I	GA	T TT	A AA u Ly	A GG	A GC	C TA	T GG	T TT	T G	TA T	AC A	GA.	AAC	ACG	T-T						
	61 A		ne	ııe	AS	p Le	u Ly	s Gl	y Al	а Ту	r Gl	y Ph	e Va	l T	yr A	ra	Asn	Thr	Dha			17	GAT	1740	)
																								580	1
174	11 GC	T T	C.I.	GAA	GT	A AT.	A AA: e Asi	r AC	r GG	A GT	A GA	ב דד	דיד יו	ים מי		~ .									
5 8	31 G1	y s	er	Glu	Va:	1 11	e Ası	Thi	r Gly	/ Val	l Ası	D Phe	- 1. - 1.o		11 A	GA (	GT.	ACA	GGA	TT	T A	AT	ACA	1800	
													- 20	u As	P A	rg (	ily	Thr	Gly	Ph	e A	sn	Thr	600	
180	1 GG	TT	TT .	AGA	AA	CO	A ATZ	اململ ا	י באי																
60	1 G1	y P	he .	Arg	Ası	Ala	A ATA	Phe	61.	, y	AC	TA:	r aa	C CI	T G	GC A	GT :	AGA	GCT	TC	A G	A.A	ATT	1860	
•							a Ile		. 610	. ASI	ini	тут	As:	n Le	u G	ly s	er 2	Arg	Ala	Se	- G	lu	Ile	620	
186																					-				
62	l Se	r Ti	17 Z	Ala	250	T	AAA Lys	CAA	GGT	TCT	CCI	GAA	CA	A AC	T CF	C G	TT 7	rgg	GAT	AAT	CA 3	-T-	AC A	1920	
					AL Y	ьys	Lys	Gln	Gly	Ser	Pro	Glu	Gl	n Th	r Hi	s V	al 7	rp.	Asp	Asr	ıIl		Arc	640	
192																									
			.T A	AT	TCT	GTT	GAT Asp	TIT	CCA	ATA	AGT	GAT	GGT	AC:	A GA	ΑА	AT C	ŤA .	CT2						
	l Ası	I PI	O A	ısn	5er	Val	Asp	Phe	Pro	Ile	Ser	Asp	Gly	Th	r Gl	u A	sn L	eu '	721.	y	AA Ta	A 1	rre	1980	
																								660	
1987	TGC	cc	A G	AT	TGG	AAT	ATA Ile	GAA	CCA	TGT	AAT	CCT	GTA	GAC	GA.	את ב	-c x	<b>,</b> ,	•						
661	Cys	Pr	o A	sp	Trp	Asn	Ile	Glu	Pro	Cys	Asn	Pro	Val	Ast	Gli	ייי איני וו דיי	-C A	AC (	JAA	GCA	CC	T A	LCA	2040	
			•														· A	sn (	эΣD	Ala	Pr	o I	hr	680	
2041		AG	C T	TC (	CTA	TCT	CCT	GTT	AAC	AAT	АТТ	ACT	J. J. J.	Cmm											
681	Ile	Se:	r P	he 1	Leu	Ser	Pro	Val	Asn	Asn	Ile	Th-	T.A.	GII	GA	A GC	T T	AT A	LAT '	TTA	CA	A G	TT	2100	
																								700	
2101	GAA Glu	GTT	L A	AT (	GCT	ACT	GAT	GCA	GDT	CCN							٠.								
701	Glu	Va]	L As	sn A	lla.	Thr	gaA	Ala	Acn	GGW.	ACT	ATT	GAT	AAT	GTA	AA A	A C	ŢŢ	AT I	ATA	GAT	גג	AC	2160	
	Glu								p	GIY	Int	116	Asp	Asn	Val	Ly	s Le	u T	yr :	lle	Asp	A:	sn	720	
2161																									
721	AAT Asn	Leu	Va	ג ו	200	Cla	MIM .	AAT	TCT . -	ACT '	TCA '	TAT .	AAA	TGG	GGC	CA	TTC	T G	AT 1	CT	CCA	AZ	AT	2220	
	Asn				-9	GIII	TIE.	Asn ,	Ser '	Thr !	Ser '	Tyr	Lys	Trp	Gly	Hi:	s Se	r A	sp s	er	Pro	As	sn.	740	
2221																									
	Thr	762	CI		TT I	AAT	GGT (	CTT	ACA (	SAA (	GGA )	ACT :	TAT	ACC	TTA	AA	4 GC	A A:	TT C	מים	<del>ت</del> م	G N		2280	
	Thr	J	GI	u L	eu /	Asn (	Gly 1	Leu 1	Chr (	Slu (	sly :	Thr :	Tyr	Thr	Leu	Lys	Al	a I	le A	la	Thr	D.A.		2280 760	
																								760	
2281 761	AAC	GAC	GG	G G	CT 1	CT ;	ACA (	SAA A	CG C	AA 1	TT A	CG 1	ITA .	ACT	GTA	מדמ									
,01	Asn	Asp	G1;	y A	las	er :	Thr C	ilu 1	hr c	in P	he T	hr I	eu '	Thr	Val	Ile	- MU	n GF	c	AA )	agt -	CC	G	2340	
																								780	
2341 781	TCT	GAG	AA:	r ro	GT C	AC 1	TT A	AT A	.CA C	CT T	יי דם	CA 1	، بلت	. مداد	~~·	•									
781	Ser	Glu	Asr	3 C)	/s A	sp F	he A	sn T	hr P	ro S	er s	er T	hr (	201	I-I'A	GAA	GA:	TT	T G	AC A	TT	AA.	A	2400	
									_		- <b>-</b>	1		JYY .	reA	Glu	Ası	Ph	e A	5p ]	le	Ly	5	800	•
2401	AAG :	TTT	TCI	· AA	C G	TT T	TT G	AG →	Th ~	~» ~	~~ ~														
					_	•	- J	1	ای م	GA T	CT G	GC G	GA (	CA :	ICI	ATT	AGT	` AA	T T	A A	AA	AC.	A :	2460	
									E4									٠							

501	. Ly	's Pi	ie S	er A	sn V	al p	he (	Glu	Let	4 G1;	y Se	r Gl	y Gl	y Pr	° 5e	r Le	u Se	r As	sn Le	≘u Ly	ys Thi	F 82
2461 821	TI	T AC	TA	TT A	AT T	GG A	AT 1	יכפ	C 2 2				_					•		٠	A AAC	
2521 841	AA:	C GG	T GT Y Va	TA CO	CT G	AT T	AT I	'AT 'yr	ATA Ile	AAT Asn	TTA Leu	AAA Lys	Pro	A AAZ D Lys	A AT	T AC	C TT	CA Gl:	G TT n Ph	T AA e Ly	A AAT s Asn	258( 860
861 2581													-16	PIO	AST	Phe	Asp	Gl	/ As	Ty	TGG Trp	2640 880
881												• • • •	361	rys	Thr	Asn	Asn	Phe	Thr	Ile	Tyr	2700 900
											-,-		vai	rnr	Pro	Ser	neA	Gln	Ile	Ser	AAA Lys	2760 920
921	ATT Ile	ACT Thr	GAT Asp	GA1	TCT Ser	AG1	TAT	T A	AT :	Phe 1	AAG ( Lys 1	CTT :	TAC Tyr	CCT Pro	7.AT Asn	CCT Pro	GCT Ala	TTA Leu	GAC Asp	GAA Glu	ACT Thr	2820 940
821	ATT	TTT	GTG	AGC	GCT	GAA	GA1	. 63	מ גג		י מידיי	·		GTG ( Val 1				GT 2				

Figure 17d(continued)

# Figure No. 180 Pyrococcus furiosus VC1(7EG1)

lea	ader	seq	nenc	:e: a	mino	aci	ds 1	-24											
				9		1	В		2	7		3	6						
5 '	AT	G AG	C AA	G AA	A AA	G TT	C GT	C. AT	C GT	<u>አ</u> ጥር	דמ די	·~ ~~			4	5		54 A CAG	,
	Me	Se	r Ly	s Ly	s Ly	s Phe	e Va	1 II	e Va	l Se	r Il	e 1.e	A AC	A AT	CCT	T TT	A GI	A CAG l Gln	;
						-						- 10	u In	I 11	e Le	n Fe	u Va	l Gln	
			6	_		72			8:	1		9	•						
	GC	AT:	A TA	T TT	r gr	A GAZ	AAC	TA	ר כשי	- ኮ አሮ	C TC	T (1)			9:	-		108 A <b>AA</b> T	
	Ala	, Ile	Ty:	r Phe	e Val	Glu	Lys	ту	r His	Th:	r Se	י הו	O GA	- AA(	3 TC	AC.	r TC	A AAT r Asn	
								_					a wal	p rys	s Sei	: Thi	Se:	r Asn	
			117			126			135	5		144							
	ACC	TC	L. TCI	r aca	CCA		CAA	ACA	ACE	, Cum	r TC	- 200	- 		153			162 FATT	
	Thr	Ser	Ser	Thr	Pro	Pro	Gln	Thr	Thr	Lei	ı Se:	The	Thr	Tare	GTT	CTC	AAC	ATT Ile	
														Lys	val	ren	Lys	; Ile	
			171			180			189			198	,						
	AGA	TAC	CCI	GAT	GAC	GGT	GAG	TGG	CCA	GGA	GCT			C D T	207			216 GAT	
	Arg	Tyr	Pro	qzA o	Asp	Gly	Glu	Trp	Pro	Gly	Ala	Pro	Ile	Lan	TVO	GAT	GGI	GAT Asp	
•							•	-						,	Ly S	Asp	GIY	Asp	
			225			234			243			252			261				
	GGG	AAC	CCA	GAA Glu	TTC	TAC	ATT	GAA	ATA	AAC	CTA	TGG	AAC	ATT		<u>አ</u> አጥ		270	
	GIÀ	Asn	Pro	Glu	Phe	Tyr	Ile	Glu	Ile	Asn	Leu	Trp	Asn	Ile	Leu	Asn	בוש	ACT	
																• • • • • • • • • • • • • • • • • • • •	~~~		
	~~ »		279			288			297			306			315		-	374	
	Clv	1.1.1.	GCT	GAG Glu	ATG	ACG	TAC	AAT	TTA	ACC	AGC	GGC	GTC	CTT		TAC	GTC	CDD	
· ·	Cly	PHE	Ala	Glu	Met	Thr	Tyr	Asn	Leu	Thr	Ser	Gly	Val	Leu	His	Tyr	Val	Gln	
																•			
	ממ־	ىلىناس	333			342			351			360			369			378	
,	Sln	Len	Acr	AAC Asn	ATT	GTC	TTG	AGG	GAT	AGA	AGT	AAT	TGG	GTG	CAT	GGA	TAC	CCC	
	,		vsh	Asn	TTE	Val :	Leu .	Arg	Asp	Arg	Ser	Asn	Trp	Val	His	Gly	Tyr	Pro	
																	-		

441 450 459 468 ATA CCA TTA CCC AGT AAA GTT TCA AAC CTA ACA GAC TTC TAT CTA ACA ATC TCC Ile Pro Leu Pro Ser Lys Val Ser Asn Leu Thr Asp Phe Tyr Leu Thr Ile Ser

405 GAA ATA TTC TAT GGA AAC AAG CCA TGG AAT GCA AAC TAC GCA ACT GAT GGC CCA Glu Ile Phe Tyr Gly Asn Lys Pro Trp Asn Ala Asn Tyr Ala Thr Asp Gly Pro

387

396

TAT AAA CTT GAG CCC AAG AAC GGC CTG CCA ATT AAC TTC GCA ATA GAA TCC TGG Tyr Lys Leu Glu Pro Lys Asn Gly Leu Pro Ile Asn Phe Ala Ile Glu Ser Trp TTA ACG AGA GAA GCT TGG AGA ACA ACA GGA ATT AAC AGC GAT GAG CAA GAA GTA Leu Thr Arg Glu Ala Trp Arg Thr Thr Gly Ile Asn Ser Asp Glu Gln Glu Val ATG ATA TGG ATT TAC TAT GAC GGA TTA CAA CCG GCT GGC TCC AAA GTT AAG GAG Met Ile Trp Ile Tyr Tyr Asp Gly Leu Gln Pro Ala Gly Ser Lys Val Lys Glu . 657 ATT GTA GTC CCA ATA ATA GTT AAC GGA ACA CCA GTA AAT GCT ACA TTT GAA GTA Ile Val Val Pro Ile Ile Val Asn Gly Thr Pro Val Asn Ala Thr Phe Glu Val TGG AAG GCA AAC ATT GGT TGG GAG TAT GTT GCA TTT AGA ATA AAG ACC CCA ATC Trp Lys Ala Asn Ile Gly Trp Glu Tyr Val Ala Phe Arg Ile Lys Thr Pro Ile AAA GAG GGA ACA GTG ACA ATT CCA TAC GGA GCA TTT ATA AGT GTT GCA GCC AAC Lys Glu Gly Thr Val Thr Ile Pro Tyr Gly Ala Phe Ile Ser Val Ala Ala Asn ATT TCA AGC TTA CCA AAT TAC ACA GAA CTT TAC TTA GAG GAC GTG GAG ATT GGA Ile Ser Ser Leu Pro Asn Tyr Thr Glu Leu Tyr Leu Glu Asp Val Glu Ile Gly ACT GAG TTT GGA ACG CCA AGC ACT ACC TCC GCC CAC CTA GAG TGG TGG ATC ACA Thr Glu Phe Gly Thr Pro Ser Thr Thr Ser Ala His Leu Glu Trp Trp Ile Thr AAC ATA ACA CTA ACT CCT CTA GAT AGA CCT CTT ATT TCC TAA 3' Asn Ile Thr Leu Thr Pro Leu Asp Arg Pro Leu Ile Ser *

Figure 18b(continued)

# INTERNATIONAL SEARCH REPORT

International application No. PCT/US97/22623

A. CLASSIFICATION OF SUBJECT MATTER						
IPC(6) :C07H 21/04; C12N 1/20, 1/14, 5/00, 9/38, 9/42; C08B 30/04 US CL :435/207, 209, 252.3, 254.11, 274, 275, 320.1, 325; 536/23.2						
	According to International Patent Classification (IPC) or to both national classification and IPC					
B. FIEL						
Minimum d	ocumentation searched (classification system followed	by classification symbols)				
U.S. :	435/207, 209, 252.3, 254.11, 274, 275, 320.1, 325;	536/23.2				
	· · · · · · · · · · · · · · · · · · ·					
Documentat	ion searched other than minimum documentation to the	extent that such documents are included	in the fields searched			
			0			
Electronic d	ata base consulted during the international search (na	me of data base and, where practicable	e, search terms used)			
Please See	Extra Sheet.					
	•					
C. DOC	UMENTS CONSIDERED TO BE RELEVANT	· · · · · · · · · · · · · · · · · · ·				
Category*	Citation of document, with indication, where ap-	propriets of the relevant page on	Relevant to claim No.			
Category	Chargon of document, with indication, where ap-	propriate, of the relevant passages	Relevant to claim 140.			
X	GRABNITZ et al. Structure of the B	-Glucosidase Gene bglA of	1-3, 5			
	Clostridium thermocellum: Sequence Ar	nalysis Reveals a Superfamily	species II			
A	of Cellulases and β-Glycosidases Includ	ing Human Lactase/Phlorizin				
	Hydrolase. Eur. J. Biochem. Septemb	per 1991, Vol. 200, No. 2,	4, 6-11			
	pages 301-309, see entire document.					
X	VOORHORST et al. Characterization		1-3, 5			
	β-Glucosidase from the Hyperthermo		species I and III			
Α	furiosus and Its Expression and Site-Dire					
	coli. J. Bacteriol. December 1995, Vo	1. 177, No. 24, pages 7105-	4, 6-11			
	7111, see entire document.					
	<del>'</del>	•				
Further documents are listed in the continuation of Box C. See patent family annex.						
* Special categories of cited documents: "T" later document published after the international filing date or priority						
*A* document defining the general state of the art which is not considered the principle or theory underlying the invention						
to be of particular relevance  E* carlier document published on or after the international filing date  "X* document of particular relevance; the claimed invention cannot be						
*L* document which may throw doubts on priority claim(s) or which is when the document is taken alone						
cited to establish the publication date of another citation or other special reason (as specified)  *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is						
"O" document referring to an oral disclosure, use, exhibition or other combined with one or more other such documents, such combination being obvious to a person skilled in the art						
*P* document published prior to the international filing date but later than *&* document member of the same patent family						
Date of the actual completion of the international search  Date of mailing of the international search report						
26 MARCH 1998 <u><b>2</b></u> 1 APR 1998						
Name and	Name and mailing address of the ISA/US  Authorized officer					
Commissioner of Patents and Trademarks						
Washingto	n, D.C. 20231	LISA J. HOBBS, PH.D.	TIN TOIN			
Facsimile N	No. (703) 305-3230	Telephone No. (703) 308-0196				

# WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



# INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: C07H 21/04, C12N 1/20, 1/14, 5/00, 9/38,  $\mathbf{A1}$ 9/42, C08B 30/04

(11) International Publication Number:

WO 98/24799

(43) International Publication Date:

11 June 1998 (11.06.98)

(21) International Application Number:

PCT/US97/22623

(22) International Filing Date:

8 December 1997 (08.12.97)

(81) Designated States: AU, CA, IL, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

(30) Priority Data:

60/056,916 08/949,026 6 December 1996 (06.12.96) US 10 October 1997 (10.10.97)

US

(71) Applicant (for all designated States except US): DIVERSA CORPORATION [US/US]; 10665 Sorrento Valley Road, San Diego, CA 92121 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): BYLINA, Edward, J. [US/US]; Apartment A-1, West Court, Andalusia, PA 19020 (US). SWANSON, Ronald, V. [US/US]; Apartment A, 309 No. Lemon Street, Media, PA 19063 (US). MATHUR, Eric, J. [US/US]; 2654 Galicia Way, Carlsbad, CA 92009 (US). LAM, David, E. [US/US]; 1518 West 249th Street, Harbor City, CA 90710 (US).

(74) Agent: HAILE, Lisa, A.; Fish & Richardson P.C., Suite 1400. 4225 Executive Square, La Jolla, CA 92037 (US).

**Published** 

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: GLYCOSIDASE ENZYMES

(57) Abstract

Thermostable glycosidase enzymes derived from various Thermococcus, Staphylothermus and Pyrococcus organisms is disclosed. The enzymes are produced from native or recombinant host cells and can be utilized in the food processing industry, pharmaceutical industry and in the textile industry, detergent industry and in the baking industry.

# FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
ΑU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
ΑZ	Azerbaijan	GB	United Kingdom	MC	Моласо	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	-	Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
ВJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR-	Mauritania	UG	Uganda
BY	Belarus	IS	lceland -	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JР	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	· LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		



### PCT

# WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: WO 98/24799 (11) International Publication Number: C07H 21/04, C12N 1/20, 1/14, 5/00, 9/38, A1 11 June 1998 (11.06.98) (43) International Publication Date: 9/42, C08B 30/04 LAM, David, E. [US/US]; 1518 West 249th Street, Harbor (21) International Application Number: PCT/US97/22623 City, CA 90710 (US). (22) International Filing Date: 8 December 1997 (08.12.97) (74) Agent: HAILE, Lisa, A.; Fish & Richardson P.C., Suite 1400, 4225 Executive Square, La Jolla, CA 92037 (US). (30) Priority Data: 60/056,916 6 December 1996 (06.12.96) US 08/949, 026 10 October 1997 (10.10.97) US (81) Designated States: AU, CA, IL, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications Published US 60/056,916 (CIP) Filed on 6 December 1996 (06.12.96) With international search report. Before the expiration of the time limit for amending the US Not furnished (CIP) 10 October 1997 (10.10.97) claims and to be republished in the event of the receipt of Filed on amendments. (71) Applicant (for all designated States except US): DIVERSA CORPORATION [US/US]; 10665 Sorrento Valley Road, San Diego, CA 92121 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): BYLINA, Edward, J. [US/US]; Apartment A-1, West Court, Andalusia, PA 19020 (US). SWANSON, Ronald, V. [US/US]; Apartment A, 309 No. Lemon Street, Media, PA 19063 (US). MATHUR, Eric,

### (54) Title: GLYCOSIDASE ENZYMES

J. [US/US]; 2654 Galicia Way, Carlsbad, CA 92009 (US).

### (57) Abstract

Thermostable glycosidase enzymes derived from various Thermococcus, Staphylothermus and Pyrococcus organisms is disclosed. The enzymes are produced from native or recombinant host cells and can be utilized in the food processing industry, pharmaceutical industry and in the textile industry, detergent industry and in the baking industry.

# FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
ΑU	Australia	GA	Gabon	LV .	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Моласо	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of Americ
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JР	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
СН	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

CORRECTED VERSION.

CORRECTED VERSION**

CORRECTED VERSION***

# WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT) (51) International Patent Classification 6: (11) International Publication Number: WO 98/24799 C07H 21/04, C12N 1/20, 1/14, 5/00, 9/38, A1 (43) International Publication Date: 9/42, C08B 30/04 11 June 1998 (11.06.98) LAM, David, E. [US/US]; 1518 West 249th Street, Harbor (21) International Application Number: PCT/US97/22623 City, CA 90710 (US). (22) International Filing Date: 8 December 1997 (08.12.97) (74) Agent: HAILE, Lisa, A.; Fish & Richardson P.C., Suite 1400, 4225 Executive Square, La Jolla, CA 92037 (US). (30) Priority Data: 60/056,916 6 December 1996 (06.12.96) US 08/949, 026 10 October 1997 (10.10.97) US (81) Designated States: AU, CA, IL, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications Published US 60/056,916 (CIP) Filed on 6 December 1996 (06.12.96) With international search report. US Not furnished (CIP) Before the expiration of the time limit for amending the Filed on 10 October 1997 (10.10.97) claims and to be republished in the event of the receipt of amendments. (71) Applicant (for all designated States except US): DIVERSA CORPORATION [US/US]; 10665 Sorrento Valley Road, San Diego, CA 92121 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): BYLINA, Edward, J.

### (54) Title: GLYCOSIDASE ENZYMES

[US/US]; Apartment A-1, West Court, Andalusia, PA 19020 (US). SWANSON, Ronald, V. [US/US]; Apartment A, 309 No. Lemon Street, Media, PA 19063 (US). MATHUR, Eric, J. [US/US]; 2654 Galicia Way, Carlsbad, CA 92009 (US).

### (57) Abstract

Thermostable glycosidase enzymes derived from various Thermococcus, Staphylothermus and Pyrococcus organisms is disclosed. The enzymes are produced from native or recombinant host cells and can be utilized in the food processing industry, pharmaceutical industry and in the textile industry, detergent industry and in the baking industry.

## FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

			•				
· AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Słovakia
ΑT	Austria	FR	France	LU	Luxembourg	SN	Senegal
ΑU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	zw	Zimbabwe
CI	Côte d'Ivoire	KР	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		•
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## **GLYCOSIDASE ENZYMES**

## BACKGROUND OF THE INVENTION

## 1. Field of the Inventions

This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, the use of such polynucleotides and polypeptides, as well as the production and isolation of such polynucleotides and polypeptides. More particularly, the polynucleotides and polypeptides of the present invention has been putatively identified as glucosidases,  $\alpha$ -galactosidases,  $\beta$ -galactosidases,  $\beta$ -mannosidases,  $\beta$ -mannosidases, and pullalanases.

## 10 2. Description of Related Art

The glycosidic bond of  $\beta$ -galactosides can be cleaved by different classes of enzymes: (i) phospho- $\beta$ -galactosidases (EC3.2.1.85) are specific for a phosphorylated substrate generated via phosphoenolpyruvate phosphotransferase system (PTS)-dependent uptake; (ii) typical  $\beta$ -galactosidases (EC 3.2.1.23), represented by the *Escherichia coli* LacZ

- enzyme, which are relatively specific for β-galactosides; and (iii) β-glucosidases (EC 3.2.1.21) such as the enzymes of Agrobacterium faecalis, Clostridium thermocellum, Pyrococcus furiosus or Sulfolobus solfataricus (Day, A.G. and Withers, S.G., (1986) Purification and characterization of a β-glucosidase from Alcaligenes faecalis. Can. J. Biochem. Cell. Biol. 64, 914-922; Kengen, S.W.M., et al. (1993) Eur. J. Biochem., 213,
- 305-312; Ait, N., Cruezet, N. and Cattaneo, J. (1982) Properties of β-glucosidase purified from Clostridium thermocellum. J. Gen. Microbiol. 128, 569-577; Grogan, D.W. (1991) Evidence that β-galactosidase of Sulfolobus solfataricus is only one of several activities of a thermostable β-D-glycodiase. Appl. Environ. Microbiol. 57, 1644-1649). Members of the latter group, although highly specific with respect to the β-anomeric configuration
- of the glycosidic linkage, often display a rather relaxed substrate specificity and hydrolyze  $\beta$ -glucosides as well as  $\beta$ -fucosides and  $\beta$ -galactosides.

Generally,  $\alpha$ -galactosidases are enzymes that catalyze the hydrolysis of galactose groups on a polysaccharide backbone or hydrolyze the cleavage of di- or oligosaccharides comprising galactose.

Generally, \(\beta\)-mannanases are enzymes that catalyze the hydrolysis of mannose groups internally on a polysaccharide backbone or hydrolyze the cleavage of di- or oligosaccaharides comprising mannose groups. \(\beta\)-mannosidases hydrolyze non-reducing, terminal mannose residues on a mannose-containing polysaccharide and the cleavage of di- or oligosaccaharides comprising mannose groups.

Guar gum is a branched galactomannan polysaccharide composed of  $\beta$ -1,4 linked 0 mannose backbone with  $\alpha$ -1,6 linked galactose side chains. The enzymes required for the degradation of guar are  $\beta$ -mannanase,  $\beta$ -mannosidase and  $\alpha$ -galactosidase.  $\beta$ -mannanase hydrolyses the mannose backbone internally and  $\beta$ -mannosidase hydrolyses non-reducing, terminal mannose residues.  $\alpha$ -galactosidase hydrolyses  $\alpha$ -linked galactose groups.

15 Galactomannan polysaccharides and the enzymes that degrade them have a variety of applications. Guar is commonly used as a thickening agent in food and is utilized in hydraulic fracturing in oil and gas recovery. Consequently, galactomannanases are industrially relevant for the degradation and modification of guar. Furthermore, a need exists for thermostable galactomannases that are active in extreme conditions associated with drilling and well stimulation.

There are other applications for these enzymes in various industries, such as in the beet sugar industry. 20-30% of the domestic U.S. sucrose consumption is sucrose from sugar beets. Raw beet sugar can contain a small amount of raffinose when the sugar beets are stored before processing and rotting begins to set in. Raffinose inhibits the crystallization of sucrose and also constitutes a hidden quantity of sucrose. Thus, there is merit to eliminating raffinose from raw beet sugar. α-Galactosidase has also been used

as a digestive aid to break down raffinose, stachyose, and verbascose in such foods as beans and other gassy foods.

 $\beta$ -galactosidases which are active and stable at high temperatures appear to be superior enzymes for the production of lactose-free dietary milk products (Chaplin, M.F. and 5 Bucke, C. (1990) In: Enzyme Technology, pp. 159-160, Cambridge University Press, Cambridge, UK). Also, several studies have demonstrated the applicability of βgalactosidases to the enzymatic synthesis of oligosaccharides via transglycosylation reactions (Nilsson, K.G.I. (1988) Enzymatic synthesis of oligosaccharides. Trends Biotechnol. 6, 156-264; Cote, G.L. and Tao, B.Y. (1990) Oligosaccharide synthesis by enzymatic transglycosylation. Glycoconjugate J. 7, 145-162). Despite the commercial potential, only a few β-galactosidases of thermophiles have been characterized so far. Two genes reported are β-galactoside-cleaving enzymes of the hyperthermophilic bacterium Thermotoga maritima, one of the most thermophilic organotrophic eubacteria described to date (Huber, R., Langworthy, T.A., König, H., Thomm, M., Woese, C.R., Sleytr, U.B. and Stetter, K.O. (1986) T. martima sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C, Arch. Microbiol. 144, 324-333) one of the most thermophilic organotrophic eubacteria described to date. The gene products have been identified as a  $\beta$ -galactosidase and a  $\beta$ -glucosidase.

Pullulanase is well known as a debranching enzyme of pullulan and starch. The enzyme hydrolyzes α-1,6-glucosidic linkages on these polymers. Starch degradation for the production or sweeteners (glucose or maltose) is a very important industrial application of this enzyme. The degradation of starch is developed in two stages. The first stage involves the liquefaction of the substrate with α-amylase, and the second stage, or saccharification stage, is performed by β-amylase with pullalanase added as a debranching enzyme, to obtain better yields.

Endoglucanases can be used in a variety of industrial applications. For instance, the endoglucanases of the present invention can hydrolyze the internal \( \beta -1,4-glycosidic \)

bonds in cellulose, which may be used for the conversion of plant biomass into fuels and chemicals. Endoglucanases also have applications in detergent formulations, the textile industry, in animal feed, in waste treatment, and in the fruit juice and brewing industry for the clarification and extraction of juices.

## **Brief Description of the Drawings**

The following drawings are illustrative of embodiments of the invention and are not meant to limit the scope of the invention as encompassed by the claims.

Figures 1a-b are the full-length DNA and corresponding deduced amino acid sequence of M11TL of the present invention. Sequencing was performed using a 378 automated DNA sequencer for all sequences of the present invention (Applied Biosystems, Inc.).

Figure 2 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of OC1/4V-33B/G.

Figure 3 is an illustration of the full-length DNA and corresponding deduced amino acid sequence of F1-12G.

Figures 4a-b are the full-length DNA and corresponding deduced amino acid sequence of 9N2-31B/G.

Figures 5a-b are the full-length DNA and corresponding deduced amino acid sequence of MSB8-6G.

15 Figure 6 is the full-length DNA and corresponding deduced amino acid sequence of AEDII12RA-18B/G.

Figures 7a-b are the full-length DNA and corresponding deduced amino acid sequence of GC74-22G.

Figures 8a-b are the full-length DNA and corresponding deduced amino acid sequence of VC1-7G1.

Figures 9a-c are the full-length DNA and corresponding deduced amino acid sequence of 37GP1.

Figures 10a-c are the full-length DNA and corresponding deduced amino acid sequence of 6GC2.

5 Figures 11a-d are the full-length DNA and corresponding deduced amino acid sequence of 6GP2.

Figures 12a-c are the full-length DNA and corresponding deduced amino acid sequence of 63GB1.

Figures 13a-b are the full-length DNA and corresponding deduced amino acid sequence of OC1/4V.

Figures 14a-e are the full-length DNA and corresponding deduced amino acid sequence of 6GP3.

Figures 15a-d are the full-length DNA and corresponding deduced amino acid sequence of *Thermotoga maritima* MSB8-6GP2.

15 Figures 16a-c are the full-length DNA and corresponding deduced amino acid sequence of *Thermotoga maritima* MSB8-6GB4.

Figures 17a-d are the full-length DNA and corresponding deduced amino acid sequence of *Banki gouldi* 37GP4.

Figures 18a-b are the full-length DNA and corresponding deduced amino acid sequence of *Pyrococcus furiosus* VC1-7EG1.

## SUMMARY OF THE INVENTION

In a preferred embodiment of the present invention, there are provided isolated nucleic acids (polynucleotides) which encode mature enzymes having the deduced amino acid sequences of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64).

In another embodiment, the invention provides a method for producing a polypeptide including culturing host cells containing the polynucleotide of Figures 1-18 and expressing from the host cell a polypeptide encoded by the polynucleotide and isolating the polypeptide.

In another embodiment, the invention provides an enzyme selected from the group consisting of an enzyme having an amino acid sequence set forth in SEQ ID NOS: 15-28 or 61-64 and an enzyme which has at least 30 consecutive amino acid residue as an enzyme having an amino acid sequence set forth in SEQ ID NOS: 15-28 or 61-64.

In yet another embodiment, the invention provides a method for generating glucose from soluble cell oligosaccharides which includes contacting a sample containing oligosaccharides with an effective amount of an enzyme selected from the group of enzymes having the amino acid sequence set forth in SEQ ID NOS: 15-28, 61-63 and 64 such that glucose is produced

The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

### **Definitions**

"Monosaccharide", as used herein, refers to a single polyhydroxy aldehyde or ketone unit.

"Oligosaccharide", as used herein, consist of short chains of monosaccharide units joined together by covalent bonds. Of these, the most abundant are the disaccharides, which have two monosaccharide units.

"Polysaccharide", as used herein, consists of long chains having many monosaccharide units.

The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

A coding sequence is "operably linked to" another coding sequence when RNA polymerase will transcribe the two coding sequences into a single mRNA, which is then translated into a single polypeptide having amino acids derived from both coding sequences. The coding sequences need not be contiguous to one another so long as the expressed sequences ultimately process to produce the desired protein.

"Recombinant" enzymes refer to enzymes produced by recombinant DNA techniques; i.e., produced from cells transformed by an exogenous DNA construct encoding the desired enzyme. "Synthetic" enzymes are those prepared by chemical synthesis.

A DNA "coding sequence of" or a "nucleotide sequence encoding" a particular enzyme, is a DNA sequence which is transcribed and translated into an enzyme when placed under the control of appropriate regulatory sequences.

# 20 <u>Detailed Description of the Invention</u>

The polynucleotides and polypeptides of the present invention have been identified as glucosidases,  $\alpha$ -galactosidases,  $\beta$ -galactosidases,  $\beta$ -mannosidases,  $\beta$ -mannanases, endoglucanases, and pullalanases as a result of their enzymatic activity.

In accordance with one aspect of the present invention, there are provided novel enzymes, as well as active fragments, analogs and derivatives thereof.

In accordance with another aspect of the present invention, there are provided isolated nucleic acid molecules encoding the enzymes of the present invention including mRNAs, cDNAs, genomic DNAs as well as active analogs and fragments of such enzymes.

In accordance with yet a further aspect of the present invention, there is provided a process for producing such polypeptides by recombinant techniques comprising culturing recombinant prokaryotic and/or eukaryotic host cells, containing a nucleic acid sequence of the present invention, under conditions promoting expression of said enzymes and subsequent recovery of said enzymes.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotides encoding such enzymes for hydrolyzing lactose to galactose and glucose for use in the food processing industry, the pharmaceutical industry, for example, to treat intolerance to lactose, as a diagnostic reporter molecule, in corn wet milling, in the fruit juice industry, in baking, in the textile industry and in the detergent industry.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes for hydrolyzing guar gum (a galactomannan polysaccharide) to remove non-reducing terminal mannose residues. Further polysaccharides such as galactomannan and the enzymes according to the invention that degrade them have a variety of applications. Guar gum is commonly used as a thickening agent in food and also is utilized in hydraulic fracturing in oil and gas recovery. Consequently, mannanases are industrially relevant for the degradation and modification of guar gums. Furthermore, a need exists for thermostable mannases that are active in extreme conditions associated with drilling and well stimulation.

In accordance with yet a further aspect of the present invention, there are also provided nucleic acid probes comprising nucleic acid molecules of sufficient length to specifically hybridize to a nucleic acid sequence of the present invention.

In accordance with yet a further aspect of the present invention, there is provided a process for utilizing such enzymes, or polynucleotides encoding such enzymes, for *in vitro* purposes related to scientific research, for example, to generate probes for identifying similar sequences which might encode similar enzymes from other organisms by using certain regions, *i.e.*, conserved sequence regions, of the nucleotide sequence.

These and other aspects of the present invention should be apparent to those skilled in the art from the teachings herein.

The polynucleotides of this invention were originally recovered from genomic gene libraries derived from the following organisms:

M11TL is a new species of *Desulfurococcus* isolated from Diamond Pool in Yellowstone National Park. The organism grows optimally at 85-88°C, pH 7.0 in a low salt medium containing yeast extract, peptone, and gelatin as substrates with a N₂/CO₂ gas phase.

OC1/4V is from the genus *Thermotoga*. The organism was isolated from Yellowstone National Park. It grows optimally at 75°C in a low salt medium with cellulose as a substrate and N₂ in gas phase.

Pyrococcus furiosus VC1 and (7EG1) is from the genus Pyrococcus. VC1 was isolated from Vulcano, Italy. It grows optimally at 100°C in a high salt medium (marine) containing elemental sulfur, yeast extract, peptone and starch as substrates and N₂ in gas phase.

Staphylothermus marinus F1 is a from the genus Staphylothermus. F1 was isolated from Vulcano, Italy. It grows optimally at 85°C, pH 6.5 in high salt medium (marine) containing elemental sulfur and yeast extract as substrates and N₂ in gas phase.

Thermococcus 9N-2 is from the genus Thermococcus 9N-2 was isolated from diffuse vent fluid in the East Pacific Rise. It is a strict anaerobe that grows optimally at 87°C.

Thermotoga maritima MSB8 and MSB8 (Clone # 6GP2 and 6GB4) is from the genus Thermotogo, and was isolated from Vulcano, Italy. MSB8 grows optimally at  $85^{\circ}$ C, pH 6.5 in a high salt medium (marine) containing starch and yeast extract as substrates and  $N_2$  in gas phase.

Thermococcus alcaliphilus AEDII12RA is from the genus Thermococcus. AEDII12RA grows optimally at 85°C, pH 9.5 in a high salt medium (marine) containing polysulfides and yeast extract as substrates and N₂ in gas phase.

Thermococcus chitonophagus GC74 is from the genus Thermococcus. GC74 grows optimally at 85°C, pH 6.0 in a high salt medium (marine) containing chitin, meat extract, elemental sulfur and yeast extract as substrates and N₂ in gas phase. AEPII 1a grows optimally at 85°C at pH 6.5 in marine medium under anaerobic conditions. It has many substrates. Bankia gouldi is from the genus Bankia.

Accordingly, the polynucleotides and enzymes encoded thereby are identified by the organism from which they were isolated, and are sometimes hereinafter referred to as "M11TL" (Figure 1 and SEQ ID NOS:1 and 15), "OC1/4V-33B/G" (Figure 2 and SEQ ID NOS:2 and 16), "F1-12G" (Figure 3 and SEQ ID NOS:3 and 17), "9N2-31B/G" (Figure 4 and SEQ ID NOS:4 and 18), "MSB8" (Figure 5 and SEQ ID NOS:5 and 19), "AEDII12RA-18B/G" (Figure 6 and SEQ ID NOS:6 and 20), "GC74-22G" (Figure 7 and SEQ ID NOS:7 and 21), "VC1-7G1" (Figure 8 and SEQ ID NOS:8 and 22), "37GP1" (Figure 9 and SEQ ID NOS: 9 and 23), "6GC2" (Figure 10 and SEQ ID NOS: 10 and

24), "6GP2" (Figure 11 and SEQ ID NOS:11 and 25), "AEPII 1a" (Figure 12 and SEQ ID NOS:12 and 26), "OC1/4V" (Figure 13 and SEQ ID NOS:13 and 27), and "6GP3" (Figure 14 and SEQ ID NOS:28), "MSB8-6GP2" (Figure 15 and SEQ ID NOS:57 and 61), "MSB8-6GB4"(Figure 16 and SEQ ID NOS:58 and 62), "VC1-7EG1"(Figure 17 and SEQ ID NOS:59 and 63), and 37GP4 (Figure 18 and SEQ ID NOS:60 and 64).

The polynucleotides and polypeptides of the present invention show identity at the nucleotide and protein level to known genes and proteins encoded thereby as shown in Table 1.

Table 1

10	Clone	Gene/Protein with	Protein	Nucleic Acid
	M11TL-29G	Closest Homology Sulfolobus sulfataricus	Identity 51%	Identity 55%
		DSM 1616/P1, β- galactosidase		
	OC1/4V-33B/G	Caldocellum saccharolyticum, β- glucosidase	52%	57%
	Staphylothermus marinus F1-12G	Bacillus polymyxa, β-galactosidase	36%	48%
15	Thermococcus 9N2- 31B/G	Sulfolobus sulfataricus ATCC 49255/MT4, β- galactosidase	51%	50%
	Thermotoga maritima MSB8-6G	Clostridium thermocellum	45%	53%
20	Thermococcus AEDII12RA-18B/G	Bacillus polymyxa, β- galactosidase	34%	48%

	· · · · · · · · · · · · · · · · · · ·		·-	
-	Thermococcus chitonophagus GC74- 22G	Sulfolobus sulfataricus ATCC 49255/MT4, β- galactosidase	46%	54%
5	Pyrococcus furiosus VC1-7G1	Sulfolobus sulfataricus/MT-4 β- galactosidase	46.4%	52.5%
•	Thermotoga maritima α-galactosidase (6GC2)	Pediococcus pentosaceaus α-galactosidase	49%	29%
10	Thermotoga maritima  B-mannanase (6GP2)	Aspergillus aculeatus mannanase	56%	37%
	AEPII 1a ß- mannosidase (63GB1)	Sulfolobus solfactaricus ß- galactosidase	78%	56%
5	OC1/4V endoglucanase (33GP1)	Clostridium thermocellum endo-1,4-ß-endoglucanase	65%	43%
	Thermotoga mariti@aldo pullalanase (6GP3)	cellum saccharolyticum α- destrom 6 glucanohydralase	72	53
	Bankia gouldi mix Endoglucanase	None available	·	
0	(37GP1)			

The polynucleotides and enzymes of the present invention show homology to each other as shown in Table 2.

5

Table 2

	Clone	Gene/Protein with  Closest Homology	Protein Identity	Nucleic Acid Identity
- 11	Staphylothermus marinus F1-12G	Thermococcus AEDII12RA-18B/G, β- galactosidase, glucosidase	55%	57%
- 11	Thermococcus 9N2- 31B/G	Thermococcus chitonophagus GC74- 22G-glucosidase`	74%	66%
	Pyrococcus furiosus VC1-7G1	Pyrococcus furiosus VC1- 7B/G β-galactosidase	46.4%	54%

All the clones identified in Tables 1 and 2 encode polypeptides which have  $\alpha$ -glycosidase 10 or  $\beta$ -glycosidase activity.

This invention, in addition to the isolated nucleic acid molecules encoding the enzymes of the present invention, also provide substantially similar sequences. Isolated nucleic acid sequences are substantially similar if: (i) they are capable of hybridizing under conditions hereinafter described, to the polynucleotides of SEQ ID NOS: 1-14 and 57-60; (ii) or they encode DNA sequences which are degenerate to the polynucleotides of SEQ ID NOS: 1-14 and 57-60. Degenerate DNA sequences encode the amino acid sequences of SEQ ID NOS:15-28 and 61-64, but have variations in the nucleotide coding sequences. As used herein, substantially similar refers to the sequences having similar identity to the sequences of the instant invention. The nucleotide sequences that are substantially the same can be identified by hybridization or by sequence comparison. Enzyme sequences that are substantially the same can be identified by one or more of the

following: proteolytic digestion, gel electrophoresis and/or microsequencing.

One means for isolating the nucleic acid molecules encoding the enzymes of the present invention is to probe a gene library with a natural or artificially designed probe using art recognized procedures (see, for example: Current Protocols in Molecular Biology, 5 Ausubel F.M. et al. (EDS.) Green Publishing Company Assoc. and John Wiley Interscience, New York, 1989, 1992). It is appreciated to one skilled in the art that the polynucleotides of SEQ ID NOS: 1-14 and 57-60 or fragments thereof (comprising at least 12 contiguous nucleotides), are particularly useful probes. Other particular useful probes for this purpose are hybridizable fragments to the sequences of SEQ ID NOS: 1-14 and 57-60 (i.e., comprising at least 12 contiguous nucleotides).

With respect to nucleic acid sequences which hybridize to specific nucleic acid sequences disclosed herein, hybridization may be carried out under conditions of reduced stringency, medium stringency or even stringent conditions. As an example of oligonucleotide hybridization, a polymer membrane containing immobilized denatured nucleic acids is first prehybridized for 30 minutes at 45°C in a solution consisting of 0.9 M NaCl, 50 mM NaH₂PO₄, pH 7.0, 5.0 mM Na₂EDTA, 0.5% SDS, 10X Denhardt's, and 0.5 mg/ml polyriboadenylic acid. Approximately 2 X 10⁷ cpm (specific activity 4-9 X 10⁸ cpm/ug) of P end-labeled oligonucleotide probe are then added to the solution. After 12-16 hours of incubation, the membrane is washed for 30 minutes at room temperature in 1X SET (150 mM NaCl, 20 mM Tris hydrochloride, pH 7.8, 1 mM Na₂EDTA) containing 0.5% SDS, followed by a 30 minute wash in fresh 1X SET at Tm 10°C for the oligonucleotide probe. The membrane is then exposed to auto-radiographic film for detection of hybridization signals.

Stringent conditions means hybridization will occur only if there is at least 90% identity, preferably at least 95% identity and most preferably at least 97% identity between the sequences. Further, it is understood that a section of a 100 bps sequence that is 95 bps in length has 95% identity with the 1090 bps sequence from which it is obtained. See J.

Sambrook et al., Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory (1989) which is hereby incorporated by reference in its entirety. Also, it is understood that a fragment of a 100 bps sequence that is 95 bps in length has 95% identity with the 100 bps sequence from which it is obtained.

- As used herein, a first DNA (RNA) sequence is at least 70% and preferably at least 80% identical to another DNA (RNA) sequence if there is at least 70% and preferably at least a 80% or 90% identity, respectively, between the bases of the first sequence and the bases of the another sequence, when properly aligned with each other, for example when aligned by BLASTN.
- "Identity" as the term is used herein, refers to a polynucleotide sequence which comprises a percentage of the same bases as a reference polynucleotide (SEQ ID NOS:1-14 and 57-60). For example, a polynucleotide which is at least 90% identical to a reference polynucleotide, has polynucleotide bases which are identical in 90% of the bases which make up the reference polynucleotide and may have different bases in 10% of the bases which comprise that polynucleotide sequence.

The present invention relates polynucleotides which differ from the reference polynucleotide such that the changes are silent changes, for example the change do not alter the amino acid sequence encoded by the polynucleotide. The present invention also relates to nucleotide changes which result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference polynucleotide. In a preferred aspect of the invention these polypeptides retain the same biological action as the polypeptide encoded by the reference polynucleotide.

It is also appreciated that such probes can be and are preferably labeled with an analytically detectable reagent to facilitate identification of the probe. Useful reagents include but are not limited to radioactivity, fluorescent dyes or enzymes capable of catalyzing the formation of a detectable product. The probes are thus useful to isolate

complementary copies of DNA from other sources or to screen such sources for related sequences.

The polynucleotides of this invention were recovered from genomic gene libraries from the organisms listed in Table 1. For example, gene libraries can be generated in the Lambda ZAP II cloning vector (Stratagene Cloning Systems). Mass excisions can be performed on these libraries to generate libraries in the pBluescript phagemid. Libraries are thus generated and excisions performed according to the protocols/methods hereinafter described.

The excision libraries are introduced into the *E. coli* strain BW14893 F'kan1A.

Expression clones are then identified using a high temperature filter assay. Expression clones encoding several glucanases and several other glycosidases are identified and repurified. The polynucleotides, and enzymes encoded thereby, of the present invention, yield the activities as described above.

The coding sequences for the enzymes of the present invention were identified by screening the genomic DNAs prepared for the clones having glucosidase or galactosidase activity.

An example of such an assay is a high temperature filter assay wherein expression clones were identified by use of high temperature filter assays using buffer Z (see recipe below) containing 1 mg/ml of the substrate 5-bromo-4-chloro-3-indolyl-β-D-glucopyranoside (XGLU) (Diagnostic Chemicals Limited or Sigma) after introducing an excision library into the *E. coli* strain BW14893 F'kan1A. Expression clones encoding XGLUases were identified and repurified from M11TL, OC1/4V, Pyrococcus furiosus VC1, Staphylothemus marinus F1, Thermococcus 9N-2, Thermotoga maritima MSB8, Thermococcus alcaliphilus AEDII12RA, and Thermococcus chitonophagus GC74.

<u>Z-buffer:</u> (referenced in Miller, J.H. (1992) A Short Course in Bacterial Genetics, p. 445.)

per liter:

5

10

15

Na₂HPO₄-7H₂O 16.1g NaH₂PO₄-7H₂O 5.5g KCl 0.75g MgSO₄-7H₂O 0.246g β-mercaptoethanol 2.7ml

Adjust pH to 7.0

<u>High Temperature Filter Assay</u>

(1) The f factor f'kan (from E. coli strain CSH118)(1) was introduced into the pho-phh-lac-strain BW14893(2). BW13893(2). The filamentous phage library was plated on the resulting strain, BW14893 F'kan. (Miller, J.H. (1992) A Short Course in Bacterial Genetics; Lee, K.S., Metcalf, et al., (1992) Evidence for two phosphonate degradative pathways in Enterobacter Aerogenes, J. Bacteriol., 174:2501-2510.

- (2) After growth on 100 mm LB plates containing 100 μg/ml ampicillin, 80 μg/ml nethicillin and 1mM IPTG, colony lifts were performed using Millipore HATF membrane filters.
- The colonies transferred to the filters were lysed with chloroform vapor in 150 mm glass petri dishes.
  - (4) The filters were transferred to 100 mm glass petri dishes containing a piece of Whatman 3MM filter paper saturated with buffer.
    - (a) when testing for galactosidase activity (XGALase), 3MM paper was saturated with Z buffer containing 1 mg/ml XGAL (ChemBridge Corporation). After transferring filter bearing lysed colonies to the glass petri dish, placed dish in oven at 80-85°C.

25

(b) when testing for glucosidase (XGLUase), 3MM paper was saturated with Z buffer containing 1 mg/ml XGLU. After transferring filter bearing lysed colonies to the glass petri dish, placed dish in oven at 80-85°C.

'Positives' were observed as blue spots on the filter membranes. Used the 5 (5) following filter rescue technique to retrieve plasmid from lysed positive colony. Used pasteur pipette (or glass capillary tube) to core blue spots on the filter membrane. Placed the small filter disk in an Eppendorf tube containing 20 µl water. Incubated the Eppendorf tube at 75°C for 5 minutes 10 followed by vortexing to elute plasmid DNA off filter. This DNA was transformed into electrocompetent E. coli cells DH10B for Thermatoga maritima MSB8-6G, Staphylothermus marinus F1-12G, Thermococcus AEDII12RA-18B/G, Thermococcus chitonophagus GC74-22G, M11Tl and OC1/4V. Electrocompetent BW14893 F'kan1A E. coli were used for 15 Thermococcus 9N2-31B/G, and Pyrococcus furiosus VC1-7G1. Repeated filter-lift assay on transformation plates to identify 'positives'. Return transformation plates to 37°C incubator after filter lift to regenerate colonies. Inoculate 3 ml LB liquid containing 100 µg/ml ampicillin with repurified positives and incubate at 37°C overnight. Isolate plasmid DNA from these 20 cultures and sequence plasmid insert. In some instances where the plates used for the initial colony lifts contained non-confluent colonies, a specific colony corresponding to a blue spot on the filter could be identified on a regenerated plate and repurified directly, instead of using the filter rescue technique.

Another example of such an assay is a variation of the high temperature filter assay wherein colony-laden filters are heat-killed at different temperatures (for example, 105°C for 20 minutes) to monitor thermostability. The 3MM paper is saturated with different buffers (i.e., 100 mM NaCl, 5 mM MgCl₂, 100 mM Tris-Cl (pH 9.5)) to determine enzyme activity under different buffer conditions.

A β-glucosidase assay may also be employed, wherein GlcpβNp is used as an artificial substrate (aryl-β-glucosidase). The increase in absorbance at 405 nm as a result of p-nitrophenol (pNp) liberation was followed on a Hitachi U-1100 spectrophotometer, equipped with a thermostatted cuvette holder. The assays may be performed at 80°C or 90°C in closed 1-ml quartz cuvette. A standard reaction mixture contains 150 mM trisodium substrate, pH 5.0 (at 80°C), and 0.95 mM pNp derivative pNp = 0.561 mM⁻¹ cm⁻¹). The reaction mixture is allowed to reach the desired temperature, after which the reaction is started by injecting an appropriate amount of enzyme (1.06 ml final volume).

1 U  $\beta$ -glucosidase activity is defined as that amount required to catalyze the formation of 1.0  $\mu$ mol pNp/min. D-cellobiose may also be used as a substrate.

An ONPG assay for  $\beta$ -galactosidase activity is described by Miller, J.H. (1992) A Short Course in Bacterial Genetics and Mill, J.H. (1992) Experiments in Molecular Genetics, the contents of which are hereby incorporated by reference in their entirety.

A quantitative fluorometric assay for β-galactosidase specific activity is described by:

Youngman P., (1987) Plasmid Vectors for Recovering and Exploiting Tn917

Transpositions in Bacillus and other Gram-Positive Bacteria. In Plasmids: A Practical approach (ed. K. Hardy) pp 79-103. IRL Press, Oxford. A description of the procedure can be found in Miller (1992) p. 75-77, the contents of which are incorporated by reference herein in their entirety.

The polynucleotides of the present invention may be in the form of DNA which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (antisense) strand. The coding sequences which encodes the mature enzymes may be identical to the coding sequences shown in Figures 1-8 (SEQ ID NOS: 1-14 and 57-60)

or may be a different coding sequence which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same mature enzymes as the DNA of Figures 1-18 (SEQ ID NOS: 1-14 and 57-60).

The polynucleotide which encodes for the mature enzyme of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) may include, but is not limited to: only the coding sequence for the mature enzyme; the coding sequence for the mature enzyme and additional coding sequence such as a leader sequence or a proprotein sequence; the coding sequence for the mature enzyme (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequence 5' and/or 3' of the coding sequence for the mature enzyme.

Thus, the term "polynucleotide encoding an enzyme (protein)" encompasses a polynucleotide which includes only coding sequence for the enzyme as well as a polynucleotide which includes additional coding and/or non-coding sequence.

The present invention further relates to variants of the hereinabove described polynucleotides which encode for fragments, analogs and derivatives of the enzymes having the deduced amino acid sequences of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64). The variant of the polynucleotide may be a naturally occurring allelic variant of the polynucleotide or a non-naturally occurring variant of the polynucleotide.

Thus, the present invention includes polynucleotides encoding the same mature enzymes as shown in Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) as well as variants of such polynucleotides which variants encode for a fragment, derivative or analog of the enzymes of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64). Such nucleotide variants include deletion variants, substitution variants and addition or insertion variants.

As hereinabove indicated, the polynucleotides may have a coding sequence which is a naturally occurring allelic variant of the coding sequences shown in Figures 1-18 (SEQ

ID NOS: 1-14 and 57-60). As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded enzyme.

Fragments of the full length gene of the present invention may be used as a hybridization probe for a cDNA or a genomic library to isolate the full length DNA and to isolate other DNAs which have a high sequence similarity to the gene or similar biological activity. Probes of this type preferably have at least 10, preferably at least 15, and even more preferably at least 30 bases and may contain, for example, at least 50 or more bases. The probe may also be used to identify a DNA clone corresponding to a full length transcript and a genomic clone or clones that contain the complete gene including regulatory and promotor regions, exons, and introns. An example of a screen comprises isolating the coding region of the gene by using the known DNA sequence to synthesize an oligonucleotide probe. Labeled oligonucleotides having a sequence complementary to that of the gene of the present invention are used to screen a library of genomic DNA to determine which members of the library the probe hybridizes to.

The present invention further relates to polynucleotides which hybridize to the hereinabove-described sequences if there is at least 70%, preferably at least 90%, and more preferably at least 95% identity between the sequences. The present invention particularly relates to polynucleotides which hybridize under stringent conditions to the hereinabove-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. The polynucleotides which hybridize to the hereinabove described polynucleotides in a preferred embodiment encode enzymes which either retain substantially the same biological function or activity as the mature enzyme encoded by the DNA of Figures 1-18 (SEQ ID NOS: 1-14 and 57-60).

Alternatively, the polynucleotide may have at least 15 bases, preferably at least 30 bases, and more preferably at least 50 bases which hybridize to any part of a polynucleotide of

the present invention and which has an identity thereto, as hereinabove described, and which may or may not retain activity. For example, such polynucleotides may be employed as probes for the polynucleotides of SEQ ID NOS: 1-14 and 57-60, for example, for recovery of the polynucleotide or as a diagnostic probe or as a PCR primer.

Thus, the present invention is directed to polynucleotides having at least a 70% identity, preferably at least 90% identity and more preferably at least a 95% identity to a polynucleotide which encodes the enzymes of SEQ ID NOS: 15-28 and 61-64 as well as fragments thereof, which fragments have at least 15 bases, preferably at least 30 bases and most preferably at least 50 bases, which fragments are at least 90% identical, preferably at least 95% identical and most preferably at least 97% identical under stringent conditions to any portion of a polynucleotide of the present invention.

The present invention further relates to enzymes which have the deduced amino acid sequences of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) as well as fragments, analogs and derivatives of such enzyme.

- The terms "fragment," "derivative" and "analog" when referring to the enzymes of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) means enzymes which retain essentially the same biological function or activity as such enzymes. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature enzyme.
- 20 The enzymes of the present invention may be a recombinant enzyme, a natural enzyme or a synthetic enzyme, preferably a recombinant enzyme.
  - The fragment, derivative or analog of the enzymes of Figures 1-18 (SEQ ID NOS: 15-28 and 61-64) may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not

25

be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature enzyme is fused with another compound, such as a compound to increase the half-life of the enzyme (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature enzyme, such as a leader or secretory sequence or a sequence which is employed for purification of the mature enzyme or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

The enzymes and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or enzyme present in a living animal is not isolated, but the same polynucleotide or enzyme, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or enzymes could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

The enzymes of the present invention include the enzymes of SEQ ID NOS: 15-28 and 61-64 (in particular the mature enzyme) as well as enzymes which have at least 70% similarity (preferably at least 70% identity) to the enzymes of SEQ ID NOS: 15-28 and 61-64 and more preferably at least 90% similarity (more preferably at least 90% identity) to the enzymes of SEQ ID NOS: 15-28 and 61-64 and still more preferably at least 95% similarity (still more preferably at least 95% identity) to the enzymes of SEQ ID NOS: 15-28 and 61-64 and also include portions of such enzymes with such portion of the enzyme generally containing at least 30 amino acids and more preferably at least 50 amino acids.

As known in the art "similarity" between two enzymes is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one enzyme to the sequence of a second enzyme.

A variant, i.e. a "fragment", "analog" or "derivative" polypeptide, and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.

Among preferred variants are those that vary from a reference by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

15 Most highly preferred are variants which retain the same biological function and activity as the reference polypeptide from which it varies.

Fragments or portions of the enzymes of the present invention may be employed for producing the corresponding full-length enzyme by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length enzymes.

20 Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of enzymes of the invention by recombinant techniques.

Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the present invention. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

The polynucleotides of the present invention may be employed for producing enzymes by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety of expression vectors for expressing an enzyme. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the <u>E. coli. lac</u> or <u>trp</u>, the phage lambda P_L promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression

vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in <u>E. coli</u>.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as <u>E. coli</u>, <u>Streptomyces</u>, <u>Bacillus subtilis</u>; fungal cells, such as yeast; insect cells such as <u>Drosophila S2</u> and <u>Spodoptera Sf9</u>; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example; Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pD10, psiX174, pBluescript II KS, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); Eukaryotic: pSV2CAT, pOG44,

pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, P_L and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the enzymes of the invention can be synthetically produced by conventional peptide synthesizers.

- 20 Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., Molecular Cloning: A Laboratory
- 25 Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Transcription of the DNA encoding the enzymes of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Examples include the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of <u>E. coli</u> and <u>S. cerevisiae</u> TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated enzyme. Optionally, the heterologous sequence can encode a fusion enzyme including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include <u>E. coli</u>, <u>Bacillus subtilis</u>, <u>Salmonella typhimurium</u> and various species within the genera Pseudomonas, Streptomyces, and Staphylococcus, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well known to those skilled in the art.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell, 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

15

The enzyme can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The enzymes of the present invention may be a naturally purified product, or a product of chemical synthetic procedures, or produced by recombinant techniques from a prokaryotic or eukaryotic host (for example, by bacterial, yeast, higher plant, insect and mammalian cells in culture). Depending upon the host employed in a recombinant production procedure, the enzymes of the present invention may be glycosylated or may be non-glycosylated. Enzymes of the invention may or may not also include an initial methionine amino acid residue.

β-galactosidase hydrolyzes lactose to galactose and glucose. Accordingly, the OC1/4V, 9N2-31B/G, AEDII12RA-18B/G and F1-12G enzymes may be employed in the food processing industry for the production of low lactose content milk and for the production of galactose or glucose from lactose contained in whey obtained in a large amount as a by-product in the production of cheese. Generally, it is desired that enzymes used in food processing, such as the aforementioned β-galactosidases, be stable at elevated temperatures to help prevent microbial contamination.

These enzymes may also be employed in the pharmaceutical industry. The enzymes are used to treat intolerance to lactose. In this case, a thermostable enzyme is desired, as well. Thermostable  $\beta$ -galactosidases also have uses in diagnostic applications, where they are employed as reporter molecules.

Glucosidases act on soluble cellooligosaccharides from the non-reducing end to give glucose as the sole product. Glucanases (endo- and exo-) act in the depolymerization of cellulose, generating more non-reducing ends (endo-glucanases, for instance, act on internal linkages yielding cellobiose, glucose and cellooligosaccharides as products). β-glucosidases are used in applications where glucose is the desired product. Accordingly, M11TL, F1-12G, GC74-22G, MSB8-6G, OC1/4V, VC1-7G1, 9N2-31B/G and AEDII12RA18B/G may be employed in a wide variety of industrial applications, including in corn wet milling for the separation of starch and gluten, in the fruit industry for clarification and equipment maintenance, in baking for viscosity reduction, in the textile industry for the processing of blue jeans, and in the detergent industry as an additive. For these and other applications, thermostable enzymes are desirable.

Antibodies generated against the enzymes corresponding to a sequence of the present invention can be obtained by direct injection of the enzymes into an animal or by administering the enzymes to an animal, preferably a nonhuman. The antibody so obtained will then bind the enzymes itself. In this manner, even a sequence encoding only a fragment of the enzymes can be used to generate antibodies binding the whole native enzymes. Such antibodies can then be used to isolate the enzyme from cells expressing that enzyme.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, 1975, Nature, 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic enzyme products of this invention. Also, transgenic mice may be used to express humanized antibodies to immunogenic enzyme products of this invention.

- Antibodies generated against the enzyme of the present invention may be used in screening for similar enzymes from other organisms and samples. Such screening techniques are known in the art, for example, one such screening assay is described in "Methods for Measuring Cellulase Activities", *Methods in enzymology*, Vol 160, pp. 87-116, which is hereby incorporated by reference in its entirety.
- The present invention will be further described with reference to the following examples; however, it is to be understood that the present invention is not limited to such examples. All parts or amounts, unless otherwise specified, are by weight.
  - In order to facilitate understanding of the following examples certain frequently occurring methods and/or terms will be described.
- "Plasmids" are designated by a lower case p preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.
- 20 "Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1 µg of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 µl of buffer solution. For the purpose of isolating DNA

fragments for plasmid construction, typically 5 to 50 µg of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a polyacrylamide gel to isolate the desired fragment.

Size separation of the cleaved fragments is performed using 8 percent polyacrylamide gel described by Goeddel, D. et al., Nucleic Acids Res., 8:4057 (1980).

"Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T., et al., Id., p. 146). Unless otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units of T4 DNA ligase ("ligase") per 0.5 µg of approximately equimolar amounts of the DNA fragments to be ligated.

Unless otherwise stated, transformation was performed as described in the method of 20 Graham, F. and Van der Eb, A., Virology, 52:456-457 (1973).

#### Example 1

# **Bacterial Expression and Purification of Glycosidase Enzymes**

DNA encoding the enzymes of the present invention, SEQ ID NOS: 1-14 and 57-60 were initially amplified from a pBluescript vector containing the DNA by the PCR technique using the primers noted herein. The amplified sequences were then inserted into the

respective PQE vector listed beneath the primer sequences, and the enzyme was expressed according to the protocols set forth herein. The 5' and 3' primer sequences for the respective genes are as follows:

#### Thermococcus AEDII12RA -18B/G

5 5' CCGAGAATTCATTAAAGAGGAGAAATTAACTATGGTGAATGCTATGATTGTC 3' (SEQ ID NO:29)

3' CGGAAGATCTTCATAGCTCCGGAAGCCCATA 5' (SEQ ID NO:30)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Blg II.

#### OC1/4V-33B/G

- 5' CCGAGAATTCATTAAAGAGGAGAAATTAACTATGATAAGAAGGTCCGATTTTCC 3' (SEQ ID NO:31)
  - 3' CGGAAGATCTTTAAGATTTTAGAAATTCCTT 5' (SEQ ID NO:32)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Bgl Π.

- 15 Thermococcus 9N2 31B/G
  - 5' CCGAGAATTCATTAAAGAGGAGAAATTAACTATGCTACCAGAAGGCTTTCTC 3' (SEQ ID NO:33)
  - 3' CGGAGGTACCTCACCCAAGTCCGAACTTCTC 5' (SEQ ID NO:34)

Vector: pQE30; and contains the following restriction enzyme sites 5' EcoRI and 3'

20 KpnI.

Staphylothermus marinus F1 - 12G

- 5' CCGAGAATTCATTAAAGAGGAGAAATTAACTATGATAAGGTTTCCTGATTAT 3' (SEQ ID NO:35)
- 3' CGGAAGATCTTTATTCGAGGTTCTTTAATCC 5' (SEQ ID NO:36)
- Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' Bgl II.

Thermococcus chitonophagus GC74 - 22G
5'CCGAGAATTCATTAAAGAGGAGAAATTAACTATGCTTCCAGGAGAACTTTCTC 3'
(SEQ ID NO:37)

3' CGGAGGATCCCTACCCCTCTCTAAGATCTC 5' (SEQ ID NO:38)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' BamHI.

#### M11TL

5' AATAATCTAGAGCATGCAATTCCCCAAAGACTTCATGATAG 3' (SEQ ID NO:39)
3' AATAAAAGCTTACTGGATCAGTGTAAGATGCT 5' (SEQ ID NO:40)
Vector: pQE70; and contains the following restriction enzyme sites 5' Sphl and 3'
Hind III.

### Thermotoga maritima MSB8-6G

5' CCGACAATTGATTAAAGAGGAGAAATTAACTATGGAAAGGATCGATGAAATT 3' (SEQ ID NO:41)
3' CGGAGGTACCTCATGGTTTGAATCTCTTCTC 5' (SEQ ID NO:42)

Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3' KpnI.

### Pyrococcus furiosus VC1 - 7G1

5' CCGACAATTGATTAAAGAGGAGAAATTAACTATGTTCCCTGAAAAGTTCCTT 3' (SEQ ID NO:43)
3' CGGAGGTACCTCATCCCCTCAGCAATTCCTC 5' (SEQ ID NO:44)
Vector: pQE12; and contains the following restriction enzyme sites 5' EcoRI and 3'
Kpn I.

# Bankia gouldi endoglucanase (37GP1)

5' AATAAGGATCCGTTTAGCGACGCTCGC 3' (SEQ ID NO:45)
3' AATAAAAGCTTCCGGGTTGTACAGCGGTAATAGGC 5' (SEQ ID NO:46)
Vector: pQE52; and contains the following restriction enzyme sites 5' Bam HI and 3' Hind III.

Thermotoga maritima α-galactosidase (6GC2)

5' TTTATTGAATTCATTAAAGAGGAGAAATTAACTATGATCTGTGTGGAAATATTCGGAAAG 3' (SEQ ID NO:47)

3' TCTATAAAGCTTTCATTCTCTCACCCTCTTCGTAGAAG 5' (SEQ 1D NO:48)

5 Vector: pQET; and contains the following restriction enzyme sites 5' EcoRI and 3' Hind III.

Thermotoga maritima \( \mathbb{B}\)-mannanase (6GP2)

5' TTTATTCAATTGATTAAAGAGGAGAAATTAACTATGGGGATTGGTGGCGACGAC 3' (SEQ ID NO:49)

10 3' TTTATTAAGCTTATCTTTTCATATTCACATACCTCC 5' (SEQ ID NO:50)

Vector: pQEt; and contains the following restriction enzyme sites 5' Hind III and 3' EcoRI.

# AEPII 1a β-mannanase (63GB1)

5' TTTATTGAATTCATTAAAGAGGAGAAATTAACTATGCTACCAGAAGAGTTCCTATGGGGC 3'

15 (SEQ ID NO:51)

3' TTTATTAAGCTTCTCATCAACGGCTATGGTCTTCATTTC 5' (SEQ ID NO:52)

Vector: pQEt; and contains the following restriction enzyme sites 5' Hind III and 3' EcoRI.

OC1/4V endoglucanase (33GP1)

20 5'

AAAAAACAATTGAATTCATTAAAGAGGAGAAATTAACTATGGTAGAAAGACACTTCAGATATGTTCT T 3' (SEQ ID NO:53)

3' TTTTTCGGATCCAATTCTTCATTTACTCTTTGCCTG 5' (SEQ ID NO:54)

Vector: pQEt; and contains the following restriction enzyme sites 5' BamHI and 3'

25 EcoRI.

Thermotoga maritima pullalanase (6GP3)

5' TTTTGGAATTCATTAAAGAGGAGAAATTAACTATGGAACTGATCATAGAAGGTTAC 3' (SEQ ID NO:55)

3' ATAAGAAGCTTTTCACTCTCTGTACAGAACGTACGC 5' (SEQ ID NO:56)

Vector: pQEt; and contains the following restriction enzyme sites 5' EcoRI and 3' Hind III.

The restriction enzyme sites indicated correspond to the restriction enzyme sites on the bacterial expression vector indicated for the respective gene (Qiagen, Inc. Chatsworth, CA). The pQE vector encodes antibiotic resistance (Amp'), a bacterial origin of replication (ori), an IPTG-regulatable promoter operator (P/O), a ribosome binding site (RBS), a 6-His tag and restriction enzyme sites.

The pQE vector was digested with the restriction enzymes indicated. The amplified sequences were ligated into the respective pQE vector and inserted in frame with the sequence encoding for the RBS. The ligation mixture was then used to transform the E. coli strain M15/pREP4 (Qiagen, Inc.) by electroporation. M15/pREP4 contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan'). Transformants were identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies were selected. Plasmid DNA was isolated and confirmed by restriction analysis. Clones containing the desired constructs were grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture was used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells were grown to an optical density 600 (O.D. 600) of between 0.4 and 0.6. IPTG ("Isopropyl-B-D-thiogalacto pyranoside") was then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression. Cells were grown an extra 3 to 4 hours. Cells were then harvested by centrifugation.

The primer sequences set out above may also be employed to isolate the target gene from the deposited material by hybridization techniques described above.

#### Example 2

# Isolation of A Selected Clone From the Deposited genomic clones

A clone is isolated directly by screening the deposited material using the

oligonucleotide primers set forth in Example 1 for the particular gene desired to be isolated. The specific oligonucleotides are synthesized using an Applied Biosystems DNA synthesizer. The oligonucleotides are labeled with ³²P- -ATP using T4 polynucleotide kinase and purified according to a standard protocol (Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring,

transform bacterial hosts which are then plated on 1.5% agar plates to the density of 20,000-50,000 pfu/150 mm plate. These plates are screened using Nylon membranes according to the standard screening protocol (Stratagene, 1993). Specifically, the Nylon membrane with denatured and fixed DNA is prehybridized in 6 x SSC, 20 mM NaH₂PO₄, 0.4%SDS, 5 x Denhardt's 500 μg/ml denatured, sonicated salmon sperm

NY, 1982). The deposited clones in the pBluescript vectors may be employed to

DNA; and 6 x SSC, 0.1% SDS. After one hour of prehybridization, the membrane is hybridized with hybridization buffer 6xSSC, 20 mM NaH₂PO₄, 0.4%SDS, 500 ug/ml denatured, sonicated salmon sperm DNA with 1x10⁶ cpm/ml ³²P-probe overnight at 42°C. The membrane is washed at 45-50°C with washing buffer 6 x SSC, 0.1% SDS for 20-30 minutes dried and exposed to Kodak X-ray film overnight. Positive clones

are isolated and purified by secondary and tertiary screening. The purified clone is sequenced to verify its identity to the primer sequence.

Once the clone is isolated, the two oligonucleotide primers corresponding to the gene of interest are used to amplify the gene from the deposited material. A polymerase chain reaction is carried out in 25 µl of reaction mixture with 0.5 ug of the DNA of the gene of interest. The reaction mixture is 1.5-5 mM MgCl₂, 0.01% (w/v) gelatin, 20 µM each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with the Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by

agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the gene of interest by subcloning and sequencing the DNA product. The ends of the newly purified genes are nucleotide sequenced to identify full length sequences. Complete sequencing of full length genes is then performed by Exonuclease III digestion or primer walking.

### Example 3

# Screening for Galactosidase Activity

Screening procedures for  $\alpha$ -galactosidase protein activity may be assayed for as follows:

Substrate plates were provided by a standard plating procedure. Dilute XL1-Blue MRF *E coli* host of (Stratagene Cloning Systems, La Jolla, CA) to O.D.₆₀₀ = 1.0 with NZY media. In 15 ml tubes, inoculate 200 μl diluted host cells with phage. Mix gently and incubate tubes at 37 °C for 15 min. Add approximately 3.5 ml LB top agarose (0.7%) containing 1mM IPTG to each tube and pour onto all NYZ plate surface. Allow to cool and incubate at 37 °C overnight. The assay plates are obtained as substrate p-Nitrophenyl α-galactosidase (Sigma) (200 mg/100 ml) (100 mM NaCl, 100 mM Potassium-Phosphate) 1% (w/v) agarose. The plaques are overlayed with nitrocellulose and incubated at 4 °C for 30 minutes whereupon the nitrocellulose is removed and overlayed onto the substrate plates. The substrate

### Example 4

# Screening of Clones for Mannanase Activity

A solid phase screening assay was utilized as a primary screening method to test clones for \( \beta \)-mannanase activity.

25 A culture solution of the Y1090-E. coli host strain (Stratagene Cloning Systems, La Jolla, CA) was diluted to O.D.₆₀₀=1.0 with NZY media. The amplified library from Thermotoga maritima lambda gtl1 library was diluted in SM (phage dilution buffer):

 $5 \times 10^7$  pfu/µl diluted 1:1000 then 1:100 to  $5 \times 10^2$  pfu/µl. Then 8 µl of phage dilution ( $5 \times 10^2$  pfu/µl) was plated in 200 µl host cells. They were then incubated in 15 ml tubes at 37 °C for 15 minutes.

Approximately 4 ml of molten, LB top agarose (0.7%) at approximately 52 °C was added to each tube and the mixture was poured onto the surface of LB agar plates. The agar plates were then incubated at 37 °C for five hours. The plates were replicated and induced with 10 mM IPTG-soaked Duralon-UVTM nylon membranes (Stratagene Cloning Systems, La Jolla, CA) overnight. The nylon membranes and plates were marked with a needle to keep their orientation and the nylon membranes were then removed and stored at 4 °C.

An Azo-galactomannan overlay was applied to the LB plates containing the lambda plaques. The overlay contains 1% agarose, 50 mM potassium-phosphate buffer pH 7, 0.4% Azocarob-galactomannan. (Megazyme, Australia). The plates were incubated at 72 °C. The Azocarob-galactomannan treated plates were observed after 4 hours then returned to incubation overnight. Putative positives were identified by clearing zones on the Azocarob-galactomannan plates. Two positive clones were observed.

The nylon membranes referred to above, which correspond to the positive clones were retrieved, oriented over the plate and the portions matching the locations of the clearing zones for positive clones were cut out. Phage was eluted from the membrane cut-out portions by soaking the individual portions in 500 µl SM (phage dilution buffer) and 25 µl CHCl₃.

## Example 5

### Screening of Clones for Mannosidase Activity

A solid phase screening assay was utilized as a primary screening method to test clones for β-mannosidase activity.

A culture solution of the Y1090-*E. coli* host strain (Stratagene Cloning Systems, La Jolla, CA) was diluted to O.D.₆₀₀=1.0 with NZY media. The amplified library from AEPII 1a lambda gtl1 library was diluted in SM (phage dilution buffer): 5 x 10⁷ pfu/μl diluted 1:1000 then 1:100 to 5 x 10² pfu/μl. Then 8 μl of phage dilution 5 (5 x 10² pfu/μl) was plated in 200 μl host cells. They were then incubated in 15 ml tubes at 37 °C for 15 minutes.

Approximately 4 ml of molten, LB top agarose (0.7%) at approximately 52 °C was added to each tube and the mixture was poured onto the surface of LB agar plates. The agar plates were then incubated at 37 °C for five hours. The plates were replicated and induced with 10 mM IPTG-soaked Duralon-UVTM nylon membranes (Stratagene Cloning Systems, La Jolla, CA) overnight. The nylon membranes and plates were marked with a needle to keep their orientation and the nylon membranes were then removed and stored at 4 °C.

A p-nitrophenyl-β-D-manno-pyranoside overlay was applied to the LB plates

15 containing the lambda plaques. The overlay contains 1% agarose, 50 mM potassiumphosphate buffer pH 7, 0.4% p-nitrophenyl-β-D-manno-pyranoside. (Megazyme,
Australia). The plates were incubated at 72 °C. The p-nitrophenyl-β-D-mannopyranoside treated plates were observed after 4 hours then returned to incubation
overnight. Putative positives were identified by clearing zones on the p-nitrophenyl
20 β-D-manno-pyranoside plates. Two positive clones were observed.

The nylon membranes referred to above, which correspond to the positive clones were retrieved, oriented over the plate and the portions matching the locations of the clearing zones for positive clones were cut out. Phage was eluted from the membrane cut-out portions by soaking the individual portions in 500  $\mu$ l SM (phage dilution buffer) and 25  $\mu$ l CHCl₂.

### Example 6

# Screening for Pullulanase Activity

Screening procedures for pullulanase protein activity may be assayed for as follows:

Substrate plates were provided by a standard plating procedure. Host cells are diluted to  $O.D._{600} = 1.0$  with NZY or appropriate media. In 15 ml tubes, inoculate 200  $\mu$ l diluted host cells with phage. Mix gently and incubate tubes at 37 °C for 15 min. Add approximately 3.5 ml LB top agarose (0.7%) is added to each tube and the mixture is plated, allowed to cool, and incubated at 37 °C for about 28 hours.

Overlays of 4.5 mls of the following substrate are poured:

10	100 ml tot	al volume
	0.5g	Red Pullulan Red (Megazyme, Australia)
	1.0g	Agarose
	5ml	Buffer (Tris-HCL pH 7.2 @ 75 °C)
	2ml	5M NaCl
15	5ml	CaCl ₂ (100mM)
	85ml	dH ₂ O

Plates are cooled at room temperature, and thenm incubated at 75°C for 2 hours. Positives are observed as showing substrate degradation.

### Example 7

20

#### Screening for Endoglucanase Activity

Screening procedures for endoglucanase protein activity may be assayed for as follows:

- The gene library is plated onto 6 LB/GelRite/0.1% CMC/NZY agar plates
   (-4,800 plaque forming units/plate) in E.coli host with LB agarose as top agarose.
   The plates are incubated at 37°C overnight.
  - 2. Plates are chilled at 4°C for one hour.
  - 3. The plates are overlayed with Duralon membranes (Stratagene) at

room temperature for one hour and the membranes are oriented and lifted off the plates and stored at 4°C.

- 4. The top agarose layer is removed and plates are incubated at 37°C for ~3 hours.
- 5. The plate surface is rinsed with NaCl.
  - 6. The plate is stained with 0.1% Congo Red for 15 minutes.
  - 7. The plate is destained with 1M NaCl.
- 8. The putative positives identified on plate are isolated from the Duralon membrane (positives are identified by clearing zones around clones). The
   10 phage is eluted from the membrane by incubating in 500μl SM + 25μl CHCl₃ to elute.
  - 9. Insert DNA is subcloned into any appropriate cloning vector and subclones are reassayed for CMCase activity using the following protocol:
  - i) Spin 1ml overnight miniprep of clone at maximum speed for 3 minutes.
- 15 ii) Decant the supernatant and use it to fill "wells" that have been made in an LB/GelRite/0.1% CMC plate.
  - iii) Incubate at 37°C for 2 hours.
  - iv) Stain with 0.1% Congo Red for 15 minutes.
  - v) Destain with 1M NaCl for 15 minutes.
  - vi) Identify positives by clearing zone around clone.

Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, within the scope of the appended claims, the invention may be practiced otherwise than as particularly described.

20

#### WHAT IS CLAIMED IS:

- 1. An isolated polynucleotide selected from the group consisting of:
  - (a) SEQ ID NOS: 1-14 and 57-60;
  - (b) SEQ ID NOS: 1-14 and 57-60, wherein T can also be U;
  - (c) polynucleotide sequences complementary to SEQ ID NOS: 1-14 and 57- 60;
  - (d) polynucleotide sequences which encode an amino acid sequence as set forth in SEQ ID NOS:15-28, and 61-64; and
  - (e) fragments of (a), (b), (c) or (d) that are at least 15 consecutive bases in length and that will selectively hybridize to DNA which encodes a polypeptide of SEQ ID NOS:15-28, and 61-64.
- 2. A vector comprising a polynucleotide of claim 1.
- 3. A host cell containing the vector of claim 2.
- 4. The method of claim 3, wherein the host cell is a eukaryotic cell.
- 5. The method of claim 3, wherein the host cell is a prokaryotic cell.
- 6. A method for producing a polypeptide comprising:
  - (a) culturing the host cells of claim 3;
  - (b) expressing from the host cell of claim 3 a polypeptide encoded by said polynucleotide; and
  - (c) isolating the polypeptide.

7. An enzyme selected from the group consisting of:

- (a) an enzyme comprising an amino acid sequence set forth in SEQ ID
   NOS: 15-28 or 61-64; and
- (b) an enzyme which comprises at least 30 consecutive amino acid residue as an enzyme of (a).
- 8. An enzyme of which at least a portion is coded for by a polynucleotide of claim 1, and which is selected from the group consisting of:
  - (a) an enzyme comprising an amino acid sequence which is at least 70% identical to an amino acid sequence selected from the group of amino acid sequences set forth in SEQ ID NOS:15-28 or 61-64; and
  - (b) an enzyme which comprises at least 30 amino acid residues to the enzyme of (a).
- 9. A method for generating glucose from soluble cell oligosaccharides comprising contacting a sample containing oligosaccharides with an effective amount of an enyzme selected from the group consisting of an enzyme having the amino acid sequence set forth in SEQ ID NOS: 15-28, 61-63 and 64 such that glucose is produced.
- 10. The method of cliam 9, wherein the sample is selected from the group consisting of dairy products, fruit juices, detergents, textiles, guar gum, animal feed, plant biomass and waste products.
- 11. The method of claim 9, wherein the oligosaccharide is selected from the group consisting of maltose, cellobiose, lactose, sucrose, raffinose, stachyose, verbascose, cellulose, starch, amylose, glycogen, disacharrides, polysacharrides and pullulan.

# M11TL GLYCOSIDASE - 29G COMPLETE GENE SEQUENCE - 9/95

	1 -1	17: A		***																				
	) 177 ) Mi	·I 1.	70: 19 70: 19	14. (3	V	۸ <b>۸</b> (:	V1L	J.f., V	n. 1	V.L.V	133	TA	.14.	יון אי	ר יו־י	W.1.	177	: 7"	r c	\A T	7-1-	٠:٨	A 1317	4.11
		,	,,,	. 1.1		yr: A	zi. i.	ite. W	•••	1	1:17	· Ty	':	1 14	•• :	; <del> 1</del>	** : •	111		10 10	1	:::	1 (1) T	. 05.
1.																								
.:	1 61	v 11	e P	rn Cl			11. Ci	AT ()		MT	VI SL	. CV.	r yra	T: 74	x; (	'A'E	TCC	UT	1; ('/	er c	νη. ۱	171	: (:Ar:	1.20
		•			, ,		111 A	els i.	,	\×n	21.1	VK!	יויד כי	יד ת	ים ע	'n 1	Trp	Va	1 11	N A	S)	۰,۰	: GAG	40
121		it: At	א נגנ	:A cr	7 CC	:A (**																		
4	l As	n Th	11 A	A AL	a GI	w La				. 1	1.771	Disc	r cara	- 62	v: v	VC.	Cic	CC	A CH	T T.	٧ı	17 2	: ለለተ	180
						7	, u	., .,		ı y	er (fire	1,116	. 472	0 C1	u A	รถ	C1 A	P.T	o (;)	y T	۱, ۱	ון ני	፡ ለአጥ ፡ ለአጥ	40
181	17	'A AA	c c	·Α ΑΛ	TCA	C C	· · · ·	, , .		, <del>-</del> -													GGC	
61	Let	u As	n G	n As	n As	n Hi	- A	L	A	1.	0A0	Lve				TT.	AAC	٧Ċ	L V.1	T A(	:v (	717	GIY	240
						,						Lys	Let		y v	<b>a</b> l.	ASN	Th	r 11	G V1	۱ ن	/a l	Cly	80
241	GT	T GA	C TC	C AC	TAC	G AT	7 77	T C	· A A	λĊ	CCA	ACT		~		~~								
8 1	Va	1 G1	u Tr	p Se	r Ar	a Il	e Pr	e Pr	o L	vs	Pro	The	. Pha	. ~	- 1		~~~	GTO	CC	TGT	'A (	:VC	AGA	300
																							_	100
301	GA	T GA	G AA	C GG	C AG	C AT	T G7	T C	cc	TA (	GAT	CTC	GAT	r ca	T A			~						
101	λs	p Gl	u As	n Gl	y Se	r Il	e Va	1 Hi	5 V	a) /	Asp	Val	Ase	. u	D 13	V .	112	U-1	CA	A AG		77	GAT	360
																								120
361	GA.	A TT	A GC	C AA	CAA	G GA	G GC	כ כז	'A A	AC (	TAT	TAC	GTA	GA	רג ג	rc s	ጉልጥ						<b>.</b>	
121	C1	u Le	u Al	a As	n Ly	s Gl	u Al	a Va	1 A	en )	lis	Tyr	Val	Gl	u Me	· ·	~-	1.4		- 10		77	CAY	420
																								140
421	, AC	/ cc.	LYC	A AA	A CI	T AT	A CT	CAA	TT	TA T	CAC	CAT	TGG	cc	c c7	rc c	CT	CTC	- TC:	-	<b>*</b> ^			400
141	yr.	G1;	y Ar	g Ly	s Le	u Il	e Le	u As	n Le	21.1	'nΓ	His	Trp	Pre	o Le	eu P	ro	Leu	Tr	l.e		nc i e	ACD	480 160
																								100
481	CC	\ ATC	TA	GTO	S AG	A AG	A AT	CC	C C	CC	'nC	AGA	CCC	ccc	TO	:A G	GC	TGG	CT	AA 1	c	A.C	GAG	540
161	Pro	) Ile	. Me	t Va	l Ar	z Ar	g Me	t Gl	y Pi	O A	sp	Arg	Ala	Pro	Se	r G	ly	TED	Lev	As:	ם מ	lu	Glu	180
			•																					
541	TCU	GTC	CI	CAC	3 77	CC	C AA	A TA	c GC	.c c	CA	TAC	ATT	CC	TG	G A	ᄊ	ATG	GGC	: GA	3 C	īλ	CCT	600
181	261	VA.	Va.	l Glu	2 Phe	e Ala	a Ly	s Ty	r Y]	.а А	la	Tyr	Ile	YTS	T	D L	ys.	Met	Gly	Gli	ı L	eu	Pro	200
601																								
201	01.1	ATC	100	AGC	ACC	: ATC	S AN	: GA	A CC	C >	АC	CTC	CII	TAI	CA	C C	XX	CCA	TAC	AT	3 7	rc	GTT	660
201	V4.1	net	111	Ser	Thi	Me	: YZI	2 C1	u Pr	O Y	sn '	۷al	Val	Tyr	. CI	u G	ln ·	Cly	Tyr	Met	: P)	)e	Val	220
661																								
221	1.45	Gly	- C1	TTC		CCC	GCC	TAG	T	GY	CT	TTC	GXA	CCI	, ec	T G	AT.	AAG	CCC	YCC	a ac	ïλ	AAT	720
	-, -	<b>U</b> 1,	<b>U</b> 13	/ Phe	PIC	PIC	, er?	TY	: Le	u 5	er	Leu	Glu	λla	A)	A A	SD :	Lys	Ala	Yr	λı	g	Asn	240
721	ATG	ATC	CAC	CCT																				
241	Met	Ile	G) r	GCT	u.	31.			. 17	7 6	AC /	AAT	ATT		_ CG	C 27	rc .	YCI	AAG	w	, co		CII	780
				Ala		VIE	AFG	VI	ı ıy		Sp /	ASD	116	Lys	Ar	g Pi	he :	Ser	Lys	Lys	PI	0	Val	260
781	GGA	CTA	ATA	TAC	CCT	446		TO	-			~~.			_~									
261	Gly	Leu	Ile	Tyr	λla	Phe	Clr	7	Ph	- 6	1	Lau	110	Clu	C1.	1 C		CA	GAA	CTX	. 77	7 (	GAT	840
				-			•	• • • •	• • • • •			56.0	LEU	Gru	61	y P:		41 E	CIA	Val	Ph	•	ASP	280
841	AAG	TIT	AAG	AGC	TCT	AAC	TTA	TAC	TA	T T	ר ז	CA	CAC	ATA	C-71		·~ ,		~~	. ~				
281	Lys	Phe	Lys	Ser	Ser	Lvs	Leu	T∨r	T~:	r Pi	10 7	thr	Acn	714	Val		/	~	COT	AGT	TC		ATC	900
						-, -		-,-	• • •						V = 2			·y s	CIY	SEL	> e	r.	116	300
901	ATC	AAT	GTI	CYY	TAC	AGG	AGA	GAT	CT	r GC	:c x	WT.	AGG	CTA	GAC	. 10		~~	ccc			٠,	->-	960
301	Ile	Asn	Val	Glu	Tyr	Arg	Arg	Asp	Lev	. Al	٠,	sn.	Ara	Leu	Ast	Tr	- I	4017	Glv	Val	~~	. 1	N/E	320
																								340
961	TAT	AGC	CCI	TTA	CTC	TAC	XXX	ATC	CTC	: GA	TG	AC .	***	CCI	ATA	AT	·c c	TG	CAC	GGG	TA	T (	CA	1020
321	Tyr	Ser	Arg	Leu	Va]	Tyr	Lys	Ile	V۵	λs	PA	sp 1	Lys	Pro	Ile	11	e L	eu	His	Glv	7∨		:1~	340
1021	TTC	CIT	TCT	ACA	CCI	CCC	CCC	ATC	AGC	: cc	CC	CT (	GAA .	AAT	CCT	. TC	Tλ	GC .	GAT	TTT	GC	3 1	cc	1080
341	Phe	Leu	Cys	Thr	Pro	CJA	Cly	Ile	Ser	. br	οÀ	la (	Slu .	Asn	Pro	СУ	s S	er.	Asp	Phe	G1	v 7	'ro	360
																								•
1081	GAC	CTC	TAT	CCT	CVY	GGA	CTC	TAC	CTA	CI	T C	TA A	<b>W</b> (	GAA	CTT	TA	CA	AC (	CGA	TAC	GC	3 0	TA	1140
361	GIU	VAI	Tyr	Pro	C) u	CIA	Leu	Tyr	Leu	Le	u L	eu I	ys (	Glu	ren	Ту	r A	sn .	Arg	Tyr	Cl	ν	al	380
141																								
141	GAC	1.10	ATC	GTC	vc.c.	CVC	WC	CCT	CTT	TC	A G	VC. N	cc /	ACK:	GAT	CC	CT	TC A	AGA	CCG	GC	T	AC .	1200
381	w2i)	reu	110	Val	The	Glu	۸sn	CIA	Va )	Şe	r A	sp s	er /	۸rq	Λsp	A1	a L	eu /	Arg	Pro	Ala	. T	yı	400
201																								
40)	CTC	uat. Val	106	CAT	177	TAC	AGC	CTA	T(;C;	*	A G	ככ נ	כדר ו	VVC.	COC	CCC	c. ¥.	TT (	CC	GTC	٨N	G	CC.	1260
1	Lens	401		nis	v.) I	Түз	Ser	Val	Trp	Ly	5 A	la A	la /	Az:n	C:14	C:15	y L	le i	, LO	Va I	Ly		ly	420
421	TAC		tri e	Trans.	veg.	1-1(;	VCV	CVI.	AAT	TA		AL; T	<b>12:</b> (	34.	(,Ve)	CCC	. 1-	μ. /	w:c:	CVC	۸۸۸	· T	Tre :	14.10
•	Tyr			, ,		14.11	7111	vab	V::::	Ty	r 1;	lu T	Th V	11	(:11	GLY	1 11	10- /	irg	(;) 11	1.y:	11	1	440

Figure la

1 488 8	Th.	(72;	(iAti	A71.	1.4.4	 . :				•		V1.1	1,1	•••	A 1	1.000	OTE. Vall	4 ****
1441	CAG GIn	TAA	14	46			•				1,11	Gln	His	1.471	.Thr	lang	11	4110

Figure 1b(Continued)

# OC1/4 GLYCOSIDASE - 33G/B COMPLETE GENE SEQUENCE - 9/95

	I ATY; ATA ACA ACA
	Het The Arg Ard GAT TIT (T'A AAA GAT TIT ATC
	I ATG ATA AGA AGG TCC GAT TTT CCA AAA GAT TTT ATC TTC GGA ACG GCT ACG GCA TAC 60  61 CAG ATT GAA GGT GCA GCA AAC CAA UND DEED LYS AND PHE ILE PHE GIY THE ALM THE ALM ALM TYE 20
	"I CAG ATT CAA COM
	Gin Tie Glu Gly Ala Ala Asu Glu Asp Gly Arg Gly Pro Ser Tie Try Asp Val Phe Ser 40
	Ala Ase Glu Asp Gly Arg Gly Pro Ser 11 TGG GAT GTC TTT TCA 120
	121 CAC ACC CCT con
	41 His The Pro Gly Lys The Leu Arn Gly Asp The Gly Asp Val Ala Cys Asp His Tyr His 60
	THE LEW ASP Gly ASP THE GLY ASP VAL ALL CAT TAT CAC 180
	THE COA TAC AND CON THE TOTAL THE TACK
	6! Arg Tyr Lys Glu Asp Ile Gin Leu Met Lys Glu Ile Gly Leu Asp Ala Tyr Arg Phe Ser 80
	241 ATC TCC TCC TCC TCT 240
	81 Ile Ser Too CCC AGA ATT ATG CCA GAT CCC ANG ANG AND
	TIP PEO ANG ITE HEL PEO AND AND AND AND AND AND CAN AND COT COM COM
	301 TAC ARC ACC TO ASS GIVE ASS THE ASS GIVE LYS GIVEN ASS DE ASS GIVEN ASS DE ASS GIVEN ASS DE ASS GIVEN
	301 TAC AAC AGA CTC GTT GAT GAG CTT TTG AAG AAT GAT ATC ATA CCA TTC GTA ACA CTC TAT 360
	101 Tyr Asn Arg Leu Val Asp Glu Leu Leu Lys Asn Asp Ile Ile Pro Phe Val Thr Leu Tyr 120
	361 CAC TOO CLO TOO COLO TO TAT 360
	121 HIS TYP ASSET THE CCC TAC GCA CTT TAT GAA AAA GCT GGA GGA GGT
	361 CAC TGG GAC TTA CCC TAC GCA CTT TAT GAA AAA GGT GGA TGG CTT AAC CCA GAT ATA GCG 121 His Trp Asp Leu Pro Tyr Ala Leu Tyr Glu Lys Gly Gly Trp Leu Asn Pro Asp Ile Ala 140 421 CTC TAT TTC AGA GCA TAC GCA LACA G
	421 CTC TAT THE ASE INC.
	421 CTC TAT TTC AGA GCA TAC GCA ACG TTT ATG TTC AAC GAA CTC GGT GAT CGT GTG AAA CAT 480  481 TGG ATT ACA CTG AAC CAA GCA TGC GCA ACG TTT ATG TTC AAC GAA CTC GGT GAT CGT GTG AAA CAT 480
-	Ale Tyr Ala Thr Phe Het Phe Asn Glu Leu Gly Add CGT GTG AAA CAT 480
	481 TGG ATT ACT TO THE 160
	481 TGG ATT ACA CTG AAC GAA CCA TGG TGT TCT TCT TCT TCC GGT TAT TAC ACG GGA GAG CAT 540  161 Trp Ile Thr Leu Asn Glu Pro Trp Cys Ser Ser Phe Ser Gly Tyr Tyr Thr Gly Glu His 180  541 GCC CCG GGT CAT CAA ALT TTD Cys
	110 ITP Cys Ser Ser Phe Ser Gly Tyr The Chu Gh GAT 540
	541: GCC CCG GGT CAT CAA AAT TTA CAA GAA GCG ATA ATC GCG GCG CAC AAC CTG TTG AGG GAA .600
	Ala Pro Gly His Gln Asn Leu Gln Gly Ala ATA ATC GCG GCG CAC AAC CTG TTG ACC CAL
	181 Ala Pro Gly His Gln Asn Leu Gln Glu Ala Ile Ile Ala Ala His Asn Leu Leu Arg Glu 200
	601 CAT GGA CAT GCC GTC CAG GCG TCC AGA GAA GAA GAA GAT AAA GAT GGG GAA GTT GGC TTA ACC 660 201 His Gly His Ala Val Gln Ala Ser Arg Glu Glu Val Lys Asp Gly Gly Val Co
	201 His Gly His Ala Val Gln Ala Ser Arg Glu Glu Val Lys Asp Gly Glu Val Gly Leu Thr 220 661 AAC GTT GTG ATG AAA ATA GAD
	661 AAC COTT COTO 1000 1000 1000 1000 1000 1000 1000 1
	661 AAC GTT GTG ATG AAA ATA GAA CCG GGC GAT GCA AAA CCC GAA AGT TTC TTG GTC GCA AGT 720  721 CTT GTT GAT AAG TTC GTT AAR GTC GTT AAR CCC GAA AGT TTC TTG GTC GCA AGT 720  721 CTT GTT GAT AAG TTC GTT AAR GTC GTT AAR CCC
	720 The Giu Pro Gly Asp Ala Lys Pro Gly Ser Phe Jan Wall
	721 CIT GIT GAT ANG TTC GIT ANT GOD THE GOD THE GEN VAI AIR SET 240
	721 CIT GIT GAT AAG TIC GIT AAT GCA TGG TCC CAT GAC CCT GIT GIT TIC GGA AAA TAT CCC 780  781 GAA GAA GCA GIT GCA CIT TIC TIC TIC TIC TIC TIC TIC TIC TIC
	781 GAA GAA GAA CON
	781 GAA GAA GCA GTT GCA CTT TAT ACG GAA AAA GCG TTG CAA GTT CTC GAT AGC GAT ATG AAT 840 261 Glu Glu Ala Val Ala Leu Tyr Thr Glu Lys Gly Leu Gln Val Leu Acc GAT AGC GAT ATG AAT 840
	261 Glu Glu Ala Val Ala Leu Tyr Thr Glu Lys Gly Leu Gln Val Leu Asp Ser Asp Met Asn 280 841 ATT ATT TCG ACT CCT ATA GREEN 280
	841 ATT ATT TOG ACT COM 280
	841 ATT ATT TCG ACT CCT ATA GAC TTC TTT GGT GTG AAT TAT TAC ACA AGA ACA CTT GTT GTT 900- 901 TTT GAT ATG AAC AAT CCT GTT GTT GTT GTT GAT GTT GTT GAT ATG AAC ATG GTT GTT GTT GTT GAT ATG AAC AAT CCT GTT GTT GTT GTT GAT ATG AAC AAT CCT GTT GTT GAT ATG AAC AAT CCT GTT GTT GTT GAT ATG AAC AAT CCT GTT GTT GTT GTT GAT ATG AAC AAT CCT GTT GTT GTT GTT GAT ATG AAC AAT CCT GTT GTT GTT GTT GAT ATG AAC AAT CCT GTT GTT GTT GTT GTT GTT GTT GTT GT
	FIG TIE ASP Phe Phe Gly Val ASN TYF TYF THE ASK ACA CIT GIT 900
	901 TTT CAT AND 110 1 300
	JOI Phe Asp Het Ash Ash Pro Leu Cla TTT TCG TAT GTT CAG GGA GAC CTT CCC ANA ASS
	Phe Asp Het Ash Ash Pro Leu Gly Phe Ser Tyr Val Gln Gly Asp Leu Pro Lys Thr Glu 320  361 ATG GGA TGG GAA ATC TAG GGG GAG
•	961 ATG GGA TGG GAA ATC TAC CCG CAG GGA TTA TTT GAT ATG CTG GTC TAT CTG AAG GAA AGA 1020
	THE GIY TEP Glu Ile Tyr Pro Gln Gly Leu Phe Asp Mer Leu Vol TAT CTG AAG GAA AGA 1020
10	121 TAT AAA CON AND 340
3	41 TYE LAW COA CIT TAT ATC ACA GAG AAC GGG ATG GCT CCA
	TAT AAA CTA CCA CTT TAT ATC ACA GAG AAC GGG ATG GCT GGA CCT GAT AAA TTG GAA AAC 1080
10	81 GGA AGA GTT CAT CAR AGA GTU ASN 360
3	61 Gly Arg Val His Asp Asn Tyr Arg Ile Glu Tyr Leu Glu Lys His Phe Glu Lys Ala Leu 380
• •	ASH TYP Arg Ile Glu Tyr Leu Glu Lys His Phe Glu AM CCA CTT 1140
114	41 GAA GCA ATC AAT GCA GAT GTT GAT TTG AAA GGT TAC TTC ATT TGG TCT TTG ATG GAT AAC 1200
,,	GIU Ala Ile ASE Ala ASP Val ASE LOU DE COT TAC TEC ATT TGG TCT TTG ATG GAT ASE
120	Glu Ala Ile Asn Ala Asp Val Asp Leu Lys Gly Tyr Phe Ile Trp Ser Leu Het Asp Asn 400
40	TTC GAA TGG GCG TGC GGA TAC TCC AAA CCT TTC GGT ATA ATC TAC GTA GAT TAC AAT ACC 1260  Phe Glu Trp Ala Cys Gly Tyr Ser Lys Arg Phe Gly Ile Ile Tyr Val Acc 7
. •	Fine GIU TEP Ala Cys Gly Tyr Ser Lys Arg Phe GIV TIC GCT ATA ATC TAC GTA GAT TAC AAT ACC 1260
126	Phe Glu TFP Ala CYR Gly Tyr Ser Lys Arg Phe Gly Ile Ile Tyr Val Asp Tyr Asn Thr 420
42	PEO LYS ATA THE ANA GAT TEA GEG ATE TEE AND CO.
	1 CCA AAA AGG ATA TTG AAA GAT TCA GCG ATG TGG TTG AAG GAA TTT CTA AAA TCT TAA 1317 Pro Lys Arg 11e Lee Lys Asp Ser Ala Met Trp Lee Lys Glo Phe Lee Lys Ser End 419
	bys of the Lou Lys Ser End 419

Figure 2

# STAPHYLOTHERMUS MARINUS GLYCOSIDASE -COMPLETE GENE SEQUENCE 9/95

1 TTC ATA ACC	
1 MAIL AND THE CCT GAT TAT TTC TRU TITE CUA AUG TITE	
1 TTG ATA AGG TIT CCT GAT TAT TTG TR: TIT GGA AGA GGT AGA TGA TGG GAG GAG ATT G 1 Met lie Arg Phe Pro Asp Tyr Phe Leu Phe Gly Thr Ala Thr Ser Ser His Gln lie G 61 GGT AAT AAC ATA TIT AAT GOD DOOR	M; 60
61 CCT AAT AND	lu 20
61 GGT AAT AAC ATA TIT AAT GAT TGG TGG GAG TGG GAG ACT AAA GGC AGG ATT AAG GTG AG 21 Gly Asn Asn Ile Phe Asn Asp Trp Trp Glu Trp Glu Thr Lys Gly Arg Ile (Arg Acc	
ASH ASH THE Phe ASH ASP TEP TEP GIU TER GIU TER GIU TER GIU TER ALL GGC AGG ATT AAG GTG AC	SA 120
121 TCG GCT AAC COLUMN TO A THE LYS VAL A	9 46
41 SEE GLY LYS ALS CYS ASH HIS TEP GLU LEU TYF LYS GLU ASP ILE GLU LEU HET ALS GL 181 CTG GGA TAT AAT GCT TAT AGG TOTAL	
ATA CYS ASH HIS Trp Glu Leu TYT LVS COLUMN ATA GAG CTT ATG GCT CA	C 180
181 CTG GGA TAT AND GOOD TO THE ALE GI	u 60
61 Leu Gly Tyr Asn Ala Tyr Arg Phe Ser Ile Glu Trp Ser Arg Ile Phe Pro Arg Lys As	•
Ala Tyr Arg Phe Ser Ile Glu TrD Ser Arg TIA TIT CCC AGA AMA GA	T 240
241 CAT ATA CAT THE PTO ATG LYS ASI	0.8
.81 His Tie ASD TOT CIN CON THE ANT AND TAT AND GAA ATA GIT ANT CON	
241 CAT ATA GAT TAT GAG TCG CTT AAT AAG TAT AAG GAA ATA GTT AAT CTA CTT AGA AAA TAG .81 His Ile Asp Tyr Glu Ser Leu Asn Lys Tyr Lys Glu Ile Val Asn Leu Leu Arg Lys Tyr .301 CGG ATA GAA CCT GTA ATC ACT	300
301 CGG ATA CAR COM	100
101 Gly 11e Glu Pro Val 11e Thr Leu His His Phe Thr Asn Pro Gln Trp Phe Het Lys 11e 361 GGT GGA TGG ACT ACG CAN GOO	
THE THE LEW HIS HIS Phe Thr ASD PTG Gln TWO DAY AND	360
361 GGT GGA TGG ACT AGG CAN CHO THE	120
361 GGT GGA TGG ACT AGG GAA GAG AAC ATA AAA TAT TTT ATA AAA TAT GTA GAA CTT ATA GCT 421 Gly Gly Trp Thr Arg Glu Glu Asn Ile Lys Tyr Phe Ile Lys Tyr Val Glu Leu Ile Ala	
and the Liu Ash Ile Lys Tyr Phe Ile Lys Tyr Val Che Lit ATA CCT	420
421 TCC GAG ATA AAA GAC GTG AAA ATA TGG ATC ACT ATT AAT GAA CCA ATA ATA TAT GTT TTA 141 Ser Glu Ile Lys Asp Val Lys Ile Trp Ile Thr Ile Asp Glu Pro Ile Tip Tra	140
141 Ser Glu Ile Lys Asp Val Lys Ile Trp Ile Thr Ile Asn Glu Pro Ile Ile Tyr Val Leu 481 CAA GGA TAT ATT TCC CCC CAA DEED TO THE TOTAL CONTROL OF THE TOTAL C	000
and the lie Trp lie Thr lie Asn Glu Pro lie lie Typ Wal In	480
481 CAA GGA TAT ATT TCC GGC GAA TGG CCA CCT GGA ATT AAA AAT TTA AAA ATA GGT GAT CAA 161 Glm Gly Tyr Ile Ser Gly Glu Trp Pro Pro Gly Ile Lys Asm Leu Lyg	160
161 Gln Gly Tyr Ile Ser Gly Glu Trp Pro Pro Gly Ile Lys Asn Leu Lys Ile Ala Asp Gln GAT CAA 541 GTA ACT AAG AAT CTT TO ALL TO GLY ILE LYS Asn Leu Lys Ile Ala Asp Gln	540
FID PID Gly Ile Lys Asn Leu Lys Ile Ala Asn Cla	540
541 GTA ACT ANG ANT CIT TITA ANA GCA COT AND THE	180
541 GTA ACT ANG ANT CIT TTA ANA GCA CAT ANT GNA GCC TAT ANT ATA CIT CAT ANA CAC GGT Val Thr Lys Asn Leu Leu Lys Ala His Asn Glu Ala Tyr Asn Ile Leu His Lys His Gly 601 ATT GTA GGC ATA GCT ANA NAC AND	600
And the Ala Tyr Asn Ile Leu His Lys His Cly	200
601 ATT GTA GGC ATA GCT ANA ANC ATG ATA GCA TTT ANA CCA GGA TCT ANT AGA GGA ANA GAC 201 Ile Val Gly Ile Ala Lys Asn Het Ile Ala Phe Lys Pro Gly Ser Asn AGA GGA ANA GAC	200
THE VAL Gly Ile Ala Lys Asn Met Ile Ala Bha CCA GGA TCT AAT AGA GGA AAA GAC	660
	220
661 ATT AAT ATT TAT CAT ANA GTC GAT ANA GCA TTC AAC TGG GGA TTT CTC AAC GGA ATA TTA 221 Ile Asn Ile Tyr His Lys Val Asp Lys Ala Phe Asn TTD Gly Pho 1	
THE TYP His Lys Val ASP Lys Ala Phe ASP TO CAN THE CTC AAC GGA ATA TTA	720
721 AGG GGA GAL GGA GGA GGA GGA GGA GGA GGA	240
Arg Gly Glu Leu Glu Thr Leu Arg Gly Lys Tyr Arg Val Glu Pro Gly Asn Ile Asp Phe 781 ATA GGC ATA AAC TAT THE COLUMN TO THE COLUMN THE	
Det Glu Thr Leu Arg Gly Lys Tyr Arg Val Glu Pro Gla ART ATT GAT TTC	780
781 ATA GGC ATA AAC THE MAD THE	260
261 Ile Gly Ile AND TOT TO TOA TAT ATT GTA AAA TAT ACT TOG AND COM	
261 Ile Gly Ile Asn Tyr Tyr Ser Ser Tyr Ile Val Lys Tyr Thr Trp Asn Pro Phe Lys Leu 841 CAT ATT AAA GTC GAL GOD TO	840
841 CAT ATT AAA GTC GAA CCA TTA GAT ACA GGT CTA TGG ACA ACT ATG GGT TAC TGC ATA TAT 281 His Ile Lys Val Glu Pro Leu Asp Thr Gly Leu Trp Thr Thr Het Gly Tac TGC ATA TAT	280
281 His Ile Lys Val Glu Pro In GAT ACA GGT CTA TGG ACA ACT ATG GGT TAG TGG	
281 His Ile Lys Val Glu Pro Leu Asp Thr Gly Leu Trp Thr Thr Het Gly Tyr Cys Ile Tyr	900
901 CCT AGA GGA ATA TAT GAA GTT GTA ATG AAA ACT CAT GAG AAA TAC GGC AAA GAA ATA ATC 301 Pro Arg Gly Ile Tyr Glu Val Val Het Lys Thr His Glu Lys Tyr Glu	300
JOI Pro Arg Gly Ile Tyr Glu Val Val Het Lys Thr His Glu Lys Tyr Gly Lys Glu Ile Ile  961 ATT ACA GAG AAC GGT GTT GGT	0.60
961 The His Glu Lys Tyr Gly Lys Glu Tia Ti-	960 320
ATT ACA GAG AAC GGT GTT GCA GTA GAA AATT	320
J21 Ile Thr Glu Asn Gly Val Ala Val Gly Ass TA AGG ATT TTA TCC ATT ATC	
J21 Ile Thr Glu Asn Gly Val Ala Val Glu Asn Ash Gly Lou Ash TTA AGG ATT TTA TCC ATT ATC AGG	1020
J21 Ile Thr Glu Asn Gly Val Ala Val Glu Asn Asp Glu Leu Arg Ile Leu Ser Ile Ile Arg	
J21 Ile Thr Glu Asn Gly Val Ala Val Glu Asn Asp Glu Leu Arg Ile Leu Ser Ile Ile Arg	1020 340
J21 Ile Thr Glu Asn Gly Val Ala Val Glu Asn Asp Glu Leu Arg Ile Leu Ser Ile Ile Arg  1021 CAC TTA CAA TAC TTA TAT AAA GCC ATG AAT GAA GGA GGA AAG GTG AAA GGA TAT TTC TAC  141 His Leu Gln Tyr Leu Tyr Lys Ala Het Asn Glu Cly Ala Leu Gla GAA GGA TAT TTC TAC	1020
J21 Ile Thr Glu Asn Gly Val Ala Val Glu Asn Asp Glu Leu Arg Ile Leu Ser Ile Ile Arg  1021 CAC TTA CAA TAC TTA TAT AAA GCC ATG AAT GAA GGA GGA AAG GTG AAA GGA TAT TTC TAC  141 His Leu Gln Tyr Leu Tyr Lys Ala Het Asn Glu Gly Ala Lys Val Lys Gly Tyr Phe Tyr  1081 TGG AGC TTC ATG GAT AND GOAT AND GOAT GAA AAT GAA GGA TAT TTC TAC	1020 340
J21 Ile Thr Glu Asn Gly Val Ala Val Glu Asn Asp Glu Leu Arg Ile Leu Ser Ile Ile Arg  1021 CAC TTA CAA TAC TTA TAT AAA GCC ATG AAT GAA GGA GGA AAG GTG AAA GGA TAT TTC TAC  141 His Leu Gln Tyr Leu Tyr Lys Ala Het Asn Glu Gly Ala Lys Val Lys Gly Tyr Phe Tyr  1081 TGG AGC TTC ATG GAT AND GOAT AND GOAT GAA AAT GAA GGA TAT TTC TAC	1020 340
J21 Ile Thr Glu Asn Gly Val Ala Val Glu Asn Asp Glu Leu Arg Ile Leu Ser Ile Ile Arg  1021 CAC TTA CAA TAC TTA TAT AAA CCC ATG AAT GAA GGA GGA AAG GTG AAA GGA TAT TTC TAC  141 His Leu Gln Tyr Leu Tyr Lys Ala Het Asn Glu Gly Ala Lys Val Lys Gly Tyr Phe Tyr  1081 TCG AGC TTC ATG GAT AAT TTT GAC TGC GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA  161 Trp Ser Phe Het Asp Asn Phe Glu Trp Asp Lys Gly Phe Asp Cla	1020 340 1080
1021 Ile Thr Glu Asn Gly Val Ala Val Glu Asn Asp Glu Leu Arg Ile Leu Ser Ile Ile Arg 1021 CAC TTA CAA TAC TTA TAT AAA CCC ATG AAT CAA CGA GCA AAG CTG AAA GGA TAT TTC TAC 1081 TGG AGC TTC ATG GAT AAT TTT GAG TGG GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA 1081 TGG AGC TTC ATG GAT AAT TTT GAG TGG GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA 1081 TGG AGC TTC ATG GAT AAT TTT GAG TGG GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA 1081 TGG AGC TTC ATG GAT AAT TTT GAG TGG GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA 1081 TAT SER Phe HEL ASP ASP PHE Glu TTP ASP Lys Gly Phe Asp Cln Arg Phe Gly Leu Val 1081 GAA CTT GAT TAT AAC AGG	1020 340 1080 360
1021 Ile Thr Glu Asn Gly Val Ala Val Glu Asn Asp Glu Leu Arg Ile Leu Ser Ile Ile Arg 1021 CAC TTA CAA TAC TTA TAT AAA CCC ATG AAT CAA CGA GCA AAG CTG AAA GGA TAT TTC TAC 1081 TGG AGC TTC ATG GAT AAT TTT GAG TGG GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA 1081 TGG AGC TTC ATG GAT AAT TTT GAG TGG GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA 1081 TGG AGC TTC ATG GAT AAT TTT GAG TGG GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA 1081 TGG AGC TTC ATG GAT AAT TTT GAG TGG GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA 1081 TAT SER Phe HEL ASP ASP PHE Glu TTP ASP Lys Gly Phe Asp Cln Arg Phe Gly Leu Val 1081 GAA CTT GAT TAT AAC AGG	1020 340 1080 360
1021 Ile Thr Glu Asn Gly Val Ala Val Glu Asn Asp Glu Leu Arg Ile Leu Ser Ile Ile Arg 1021 CAC TTA CAA TAC TTA TAT AAA GCC ATG AAT GAA CGA GGA AAG GTG AAA GGA TAT TTC TAC 141 His Leu Gln Tyr Leu Tyr Lys Ala Het Asn Glu Cly Ala Lys Val Lys Gly Tyr Phe Tyr 161 TCG AGC TTC ATG GAT AAT TTT GAG TGG GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA 163 TCP Ser Phe Het Asp Asn Phe Glu Trp Asp Lys Gly Phe Asn Gln Arg Phe Gly Leu Val 164 GAA CTT GAT TAT AAG ACT TTT GAG AGA AAA CCT AGA AAA AGC GCA TAT GTA TAT AGT CAA 165 Glu Val Asp Tyr Lys Thr Phe Glu Arg Lys Pro Arg Lys Ser Ala TTC Heat GTA TAT AGT CAA	1020 340 1080 360
J21 Ile Thr Glu Asn Gly Val Ala Val Glu Asn AST GAA TTA AGG ATT TTA TCC ATT ATC AGG  1021 CAC TTA CAA TAC TTA TAT AAA GCC ATG AAT GAA GGA GGA AAG GTG AAA GGA TAT TTC TAC  141 His Leu Gln Tyr Leu Tyr Lys Ala Het Asn Glu Gly Ala Lys Val Lys Gly Tyr Phe Tyr  161 TGG AGC TTC ATG GAT AAT TTT GAG TGG GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA  161 Trp Ser Phe Het Asp Asn Phe Glu Trp Asp Lys Gly Phe Asn Gln Arg Phe Gly Leu Val  161 GAA CTT GAT TAT AAG ACT TTT GAG AGA AAA CCT AGA AAA AGC GCA TAT GTA TAT AGT CAA  163 Glu Val Asp Tyr Lys Thr Phe Glu Arg Lys Pro Arg Lys Ser Ala Tyr Val Tyr Ser Gln  160 ATA GGA CGT ACC AND ACC	1020 340 1080 360 1140 180
J21 Ile Thr Glu Asn Gly Val Ala Val Glu Asn AST GAA TTA AGG ATT TTA TCC ATT ATC AGG  1021 CAC TTA CAA TAC TTA TAT AAA GCC ATG AAT GAA GGA GGA AAG GTG AAA GGA TAT TTC TAC  141 His Leu Gln Tyr Leu Tyr Lys Ala Het Asn Glu Gly Ala Lys Val Lys Gly Tyr Phe Tyr  161 TGG AGC TTC ATG GAT AAT TTT GAG TGG GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA  161 Trp Ser Phe Het Asp Asn Phe Glu Trp Asp Lys Gly Phe Asn Gln Arg Phe Gly Leu Val  161 GAA CTT GAT TAT AAG ACT TTT GAG AGA AAA CCT AGA AAA AGC GCA TAT GTA TAT AGT CAA  163 Glu Val Asp Tyr Lys Thr Phe Glu Arg Lys Pro Arg Lys Ser Ala Tyr Val Tyr Ser Gln  160 ATA GGA CGT ACC AND ACC	1020 340 1080 360 1140 180
1021 CAC TTA CAA TAC TTA TAT AAA CCC ATG AAT GAA GGA GGA AAG GTG AAA GGA TAT TTC TCC ATT TTC TAC 1021 CAC TTA CAA TAC TTA TAT AAA GCC ATG AAT GAA GGA GGA AAG GTG AAA GGA TAT TTC TAC 141 His Leu Gln Tyr Leu Tyr Lys Ala Het Asn Glu Cly Ala Lys Val Lys Gly Tyr Phe Tyr 161 TCG AGC TTC ATG GAT AAT TTT GAG TGG GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA 161 Trp Ser Phe Het Asp Asn Phe Glu Trp Asp Lys Gly Phe Asn Gln Arg Phe Gly Leu Val 161 GAA CTT GAT TAT AAG ACT TTT GAG AGA AAA CCT AGA AAA AGC CCA TAT GTA TAT ACT CAA 161 Glu Val Asp Tyr Lys Thr Phe Glu Arg Lys Pro Arg Lys Ser Ala Tyr Val Tyr Ser Gin 162 ATA GGA CGT ACC AAG ACT ATA AGT GAT GAA TAC CTA GAA AAA TAT GGA TTA AAG AAC CTC 163 ATA GGA CGT ACC AAG ACT ATA AGT GAT GAA TAC CTA GAA AAA TAT GGA TTA AAG AAC CTC 164 Ala Arg Thr Lys Thr Ile Ser Asp Glu Tyr Leu Glu Lys Tyr Gly Leu Lys Asp Leu 165 ASP Leu Cl	1020 340 1080 360 1140 80
1221 Ile Thr Glu Asn Gly Val Ala Val Glu Asn ASP Glu Leu Arg Ile Leu Ser Ile Ile Arg 1021 CAC TTA CAA TAC TTA TAT AMA CCC ATG AAT CAA CGA GCA AAG CTG AAA GGA TAT TTC TAC 1041 His Leu Gln Tyr Leu Tyr Lys Ala Het Asn Glu Cly Ala Lys Val Lys Gly Tyr Phe Tyr 1061 TGG AGC TTC ATG GAT AAT TTT GAC TGC GAT AAA GGA TTT AAC CAA AGG TTC GGA CTA GTA 1071 TTP Ser Phe Het Asp Asn Phe Glu Trp Asp Lys Gly Phe Asn Gln Arg Phe Gly Leu Val 1072 GAA CTT GAT TAT AAG ACT TTT GAG AGA AAA CCT AGA AAA AGC CCA TAT GTA TAT AGT CAA 1073 Ile Ala CTA GAT ACC AAG ACT ATA AGT GAT GAA TAC CTA GAA AAA TAT GGA TTA AAG AAC CTC 1074 ATA GCA CGT ACC AAG ACT ATA AGT GAT GAA TAC CTA GAA AAA TAT GGA TTA AAG AAC CTC 1075 Ile Ala Arg Thr Lys Thr Ile Ser Asp Glu Tyr Leu Glu Lys Tyr Clau	1020 340 1080 360 1140 80 200 00

Figure 3

## Thermococcus 9N2 Glycosidase -318/G Complete gene bequence 9/95

												9en	c PC	i) went	ce 9	/ 75							
			TG	CTI	, נט	' cn	A GCC	717	CT(	TO	2 GC:											ccc	
		1 🕨	ec	Lau	Pro	Clu	3 Gly	/ Phe	Let	Tre	5 G)			LAN	י דכנ	: 600	111	CAG	TTC	GAC	. ATC	G CCC	60
											<b>-</b>	,	. 751	CIN	5ez	csh	Phe	Cln	Phe	Glu	Net	alv	20
	7																						
	•		•P	ry.	Leu	Arg	, YLC	A A ETT	Ile	Art	) P-0	A ===	70	CAC	TCC	100	we	TCC	CTC	ACC	CAT		120
																							40
	121	T	1.0	YYC	ATA	AAG	ACC	cu	CTC	UTC	. ACC											Pro Tac	40
	• 1	. Р	he	yen	Ile	-ye	γrg	Clu	-eu	Val	***	- Cl.	CAC	(70	CCC	CYC	CYC	CCC	ATA	ALC:	,ic	T2 ~	
																							180
																							60
	61	C1	Lu	Leu	TYI	Glu	Lys	Asp	K) =	1	7.00	GCC	المناة	CYC	CIC	CCL	CTG	λXC	CII	TAC	¥G:	3 7~7	2
																							240
	341																						80
	81	CJ	Y.	Ile	Glu	Tap	Ser	AFR	7:-	411	-		CCY	YCE	766	TTT Pbe	GTG	CAG.	GTT	GNC			
																							300
	301	CC	G	C	λGC	TAC	GGA	~~	~~~							JAA (				-Cp	ATT	e10	100
	101	۸r	g )	LEP	Ser	777	Giv	****	010	,MG	GAC	G LC	٧٨٨	ATC	GAT	JAA I	GAC :	ACG	نہے	C			
																							360
	361																						120
	121	λε	P (	24	11-	λla	) ·	W.7		CXC	ATA	CCC	TAC	TAC (	CCC	CGC (	577 I	***	E) C				
•								nis .	CIR	CIA	I Lee	Aia '	LAI.	Tyr I	Arg .	Are t	741 1		21		cre	ACC	420
	421																						140
	141	Clu	ı L	<b>44</b> (	ilv	Pho 1	ive i		ATC (	CIC	YYC	cic:	MC (	CAC 1	TC :	ACC ( The I			<b>-</b>	<b>~</b>			
																							480
																							160
	161	وعد	P	:	1 - 1	71 a 1		NG C	iac j	uc (	cc: ,	ג סדם	icc )	uc c	CT )	K SOL	TT G	CC *	~- <i>.</i>				
																							540
																							180
	181	Glu	5	r V	'A 1 1	/10 C	<b>~</b> 1	מבינ	icc y	UG 1	PAC (	300	CG 1	אב א	TC 0	CC Y	ar e	~ ~	~ ~				
																							600
																							200
	201	Val	)	א פו		~~~	X	T	TC A	אכ פ	عمر	X 523.	20 0	3C C	TT 0	TG C	ac ~	<u> </u>					
																							660
6	61	CCC	73	C 70	~ ~	cc -												. u	TY T	אב די	en y	la	220
2	71	Pzo	7~	Y 5	AY G	ייייייייייייייייייייייייייייייייייייי	Tr c	ce =	CC C	<u> </u>	II Y	TG A	ת מ	<b>==</b> =	<b>4</b> 6	== == === ============================	- A	.c. ~					
				-	· ·	*3 F	ne P	ro P	ro G	ly v	al m	6C A	5E P:	ro G1	נג עו	F Y	La Ta			<b></b>	~ c	ic	720
7																							240
2	41 .	λen	¥e.	t I	- 1	~ W	) - U	45 G	בא כז	<i>7</i> 0 07	CC T	rc w	NG AT	דג כז	ע ע	/3 Ly	- T-	- 6				_	
																							780
																							260 .
2	51 .	Ala	λs	2 20	·c >-			5C TC	:C @	VO .CC	عد <del>ه</del>	VG G7	:C 03	K AT	A AT	* TX	c		~ .~				
																							840
																							280
21	31 )	NIa .	7/2	. 2-	0 7	~ ``	2 70	בא	c (2)	C CC	× ×	c cu	ccr	cu	x 60	7 GC	A Car		~ ~.			_	
																							00
																							100
30	1 5	he I	2:5	Set	- 63		. T	. II	د هن	C 6C	TA A	CA	CAA	G GCC	ב אא	CTC	2 33/		~ ~~				
																							60
																							20
32	1 6	1v 0	lu	The		- 11-1		r (23.		CV.	1 52	CAC	i occ	. AAC	: cx	TER		COL					
	_				. P.114	• VA.	Ly	Va.	L Ars	; mi	s Lei	وعددان	7 615	/ AEE	A Aur	7	T1-		. 07	r aac	TAC	2 1	020
102	1 T	AC 1		AC1	~								•					GTA	VA.	Asz	1 37:	. 3	40 .
34	1 7	יי די	74	2-0		1 637		: YCC	TAT	י דכי	GA(	: 000	AAC	TTC	con	AGC		~~~					
																							080
108		TC C											•				116	PEG	Leu	Ila	Ser	: 3	60
361	L P)	10 A		C1.	U-1	CAC		TAC	CCC	TAC	: GCC	TCC	AGG	CCC	CCC	ACT				_	_		
		^	- ¥	01 Y	A# 1	M) a	Arm	TYE	CJA	. בגז	. Y: º	Cyx	Λrg	Pro	Glv	ACT Ser	4e-	100	. GCC	CYC	CCY	1:	140
1141		G (1		ويت									•	_	3			Sel	WIW	Arp	Cly	31	90
381	Ar	G P	 	U. I	AUC	UAC	ATC	CCC	TGG	CAG	ATC	TAT	CCC	GAG	CGG	ATC	TAC	C> C	-				
				441	ocr	Asp	Ile	Cly	TIP	Cl.	11+	Tyr	Pro	Clu	Glv	ATC 11.	***		1.00	ATA	YEY	12	100
1301	G.		٠,-											- • •	7	-14	·YZ	AED.	Ser	Ile	۾ تي	40	0
401	61	וגוו		~~	***	TAC	CCC	CTC	CCC	44.5	TAC	GTC	ACC	CAL	AAC	CCY	371						
	-1	- ^1	•	~=71	TAR	17.1	Cly	Va I	Pro	Val	Tyr	Val	Thr	G) 11	Acr	CJA CCY	MIA.	CCC	CAT	TCA	XCT		60
1261	4.0	C 2"	۰.	~~~							-	•				~1 Y	110	YIT	ALD	Ser	Thr	42	0
421	_	- ~		-TG	CCC	~-	*																
•		//	•	-æu	∧z g	PIO	IAL	Tyr	Leu	ALa	Ser	Hip	Val	Ala	Lva	ATT Ile	C1		ece	TAC	CYC	1)	
					•											44.0	U.U.	CIU	Ala	TYE	01 <i>n</i>	44	0

Figure 4a

1321 441 -381	ccc	CC:	P +1			_															
441	Ala	GIV			GTO	ACC	C-70	TAC	: CT	* TAC											
			• • •	. val	, AP	yrc	Gly	. 772	Len	•		GCC	. 52.0	ACC	GAC	: AAC	TAC	CAC			
461 1441 481	Leu	Cly	Ph.		ATC	YCC	110	600	CTC	TAT	222	C-74-					.,-		***	ATA	460
		•		~20	Het	٧rg	Phe	CIA	Luis	Tyr	Lve	Val	CAT	CJC	- ATA	ACC	MC	CAG	101		
																					1440
481	Pro	Arg	Clu	C)	ACC	CTA	ANG	CII	TAT	AGC	ccc	177	-						~~ 4	Thr AAC	480
				41U	262	Val	Lys	Val'	?yr	λεα	Giv	71	CTC	CYC	WC	WC	CCA	OTC:	MOC	110	
													AGI	Clu	yes	λan	Gly	Val	Ser	AAC .	
150; 501	Clu	11.	Ara	Clar	***	:10	CCV	CIT	GGG	TOA	15	30					-		•	Cys	500
-				oru	~).=	Pha	Cly	Leu	Glv	End											

Figure 4b(Continued)

20

120

40

ATG GAA AGG ATC GAT GAA ATT CTC TCT CAG TTA ACT ACA GAG GAA AAG Mei Gio Arg He GTG AAG CTC Giu lie Leu Ser Gin CIT Leu The The Civ Ġш 1.75 1.cu Val 61 סדס ססס הוד טהד כדד CCA GGA CTT TTT GUG AAC CCA CAT TCC AGA CTG Val Gly 1.cu CCG CCT Gly Leu Phe CCC CCT City Pro Asn His Αıχ Ser Val ۸ta GIV ۸la GGA GAA ACA CAT CCC GTT CCA AGA CTT GGA ATT CCT GCG 121 TIT Gly Glu The His כדנ כדנ CCA Val Pro Arg Leu. Pro GAT CCT 180 Cly He Pro Phe Ala Λb Asp Gly Pro 60 GCA GGA CTC AGA ATA AAT CCC ACA AGG GAA AAC GAT GAA AAC ACT TAC Ala Gly Leu Arg He Asn Pro Thr Arg Glu TAC ACG ACG GCA 240 Asp Asp Ciu Thr Thr TIT CCC GTT GAA ATC ATG CTC GCT TCT ACC TGG AAC AGA GAC CTT Ala 80 CTG GAA Val Glu lic GAA Met Leu Ala CTG Scr ` Thr GGA 300 Τmp Asa Arg Aso Leu Giv Glu اه۷ Gly 100 AAA GCC ATG GGA GAA GAA GTT AGG GAA TAC GGT GTC GAT GTG CTT 301 101 CTT Met Gly Glu Glu GCA Val Arg Glu Tyr. CCT GCG ATG 360 Gly Val Αsp Val Ala Pro Ala Met ANC ATT CAC AGA AAC CCT CTT TGT GGA AGG AAT TTC GAG TAC TAC TCA 120 361 121 His Arg Asn GAA GAT CCT Leu Cys Gly Arg Asn Phe GTC 420 Clu Tyr Tyr Ser Giv Pro ٧al CIT TOO GOT GAA ATG GOT TOA GOO TIT GTO AAG GGA GIT CAA TOT 140 421 141 Gly Glu Met Ala CAA GGG Ser Ala Phe CTG GGA GCC 480 Val Lys Gly Val Gin Gln Ciy Val Giy Ala TOC ATA AMA CAC TITT GTC GCG AAC AAC CAG GAA ACG AAC AGG ATG 160 481 161 . CTA CTG Val GAC Ala Azo Azo ACG ATC 540 G)n Glu Thr Asa Arg Met Val Val Asp He GTG TCC GAG CGA GCC CTC AGA GAA ATA TAT CTG AAA GGT TTT 180 541 181 Glu Arg Ala GAA Leu ATT CCT CTC Arg Glu lie 600 Tyr Leu Lys Gly Phe lle Ala Lys Val Lys GCA AGA CCC TGG ACC GTG ATG AGC GCT TAC AAC AAA CTG AAT GGA AAA 200 601 201 Val Met Ser TAC TOT TCA CAG 660 Αla Asn Lys Tyr Leu Gly Lys Cys Ser Gin 220 661 MC GAA TGG CTT TTG ANG ANG CIT CTC AGG GAN GAN TGG GGA 221 Asn Glu 111 GGC Trp Leu Leu GGT TTC Lys Val Leu ATG 720 Arg Glu Clu Trp Gly Gly Met AGE GAC TGG TAE GCG GGA GAE AAC CCT GTA GAA CAG CTE AAG GCC 240 721 241 Trp Tyr Gly Asp Asn MAC GAT ATG ATC 780 Pro Val Gin Lev Lys Gly Asn McI GGG AAA GCG TAT CAG GTG AAC ACA GAA AGA AGA GAT GAA ATA ATG CCT Met Pro GAA Tyr Gin Val Asn GAA ATC 840 Thr Glu Arg Arg Giv Glu Hc 280 GAG GCG TTG AAG GAG GGA AAA TTG AGT GAG GAG GTT CTC GAT GAG TGT Gly Lys Leu Ser CTG AGA AAC 900 Glu Leu Cys Val Arg Asn CTC AAA GIT CTT GTG AAC GCG CCT TCC TTC AAA GGG TAC AGG TAC 901 301 Val TCA Asn Ala Pro Ser AAC AAG CCG 960 Phe Lys Cly Tyr Arg Tyr Ser Asn Pro CTC GAA TCT CAC, GCG GAA GTC GCC TAC GAA GCA GGT Lys 320 A₅p 961 321 GCG GAG GGT Ser His Αla GTI Val Ala Glu CTC CTT CTT GAG 1020 Tyr Ciu Gly Ala Glu Cly Val Val AAC AAC GGT GTT CTT CCG TTC GAT GAA AAT ACC CAT Lev Clu 1021 341 CTC Gly Val CCC टाट गा Leu Phe Axp Pro GGC CCT Giu Axa 1080 The His Val Val Pho Cly 360 ATC GAA ACA ATA AAG GGA GGA ACG GGA AGT GGA GAC 1081 Giy Cin 361 ACC CAT Thr lie CCG AGA 1.yx Gly Gly Thr TAC ACG ATC TCT 1140 Cly Cly Aqu The Pru ATC CTT GAA GGC ATA AAA GAA AGA AAC ATG AAG ITC Tyr Thr He Ser. 380 Glu Gly He Lys Glu Arg Asn GAC GAA GAA CTC CCT TCC TAT ACT 1200 Mei Lys Pho Амр Glu Clu Ala 400 Sei The Tyr

Figure:.5a

4	01- 201	GAG Glu	GAC Glu	; TA -Ty	C AT	ΓΑ Α. L)	AA A	ΛG A ys b	TG A	GĄ G	^^ ^ lu T	CA (	AG (	iAA 1	TAT Tyr	AAA Lys	ercre Pro	∵ AG Aig					roc	1240
	261 21	GGA Gly	ACC Thr	СТ Va	C AT	A A/ Ly	A C	CG A	۸۸ C پتاند	TC C	CA G. 11 G	AG A	AT T	TC C	ਹ <b>਼</b>	TCA Sei			A GA	G AT	_		'ም 'ላል	1320
13 44	121	CCT Pro	CCA Pro	AA Lys	G AA	A A A	C G/	NT G	TT G(	A 67	n 6	TT C	TG A	TC A	.c. ,		ATC	Lys TCC	GIL GG		L ₁	_	ys AC	440
	81 (	SAC .	AGA	AA	ه دد	G CT	G AA	· ·		· ·	· va	11 Y	F1. 19	: 5	CT /	Ary	lic	Ser	C)-				yr C	1380 460
							•		, , -4		,	· u	:u 5¢	r A	SD /	Aso Oak	Cita	Len	Glu Glu				ÁA ⁄s	1440 480
									, ,,4		i Gij	r Ly	٤ Ly	5 V:	ı V	/al	Val	1 en	1			C G(	SA.	1500 500
						•					, ~Sp		. AF	. As	D G	lv '	i le	1	1	CTC Val		G CA	G	1560
							-	-			~~	~~	. A 51	اعا	ı V	ai (	GGA Giy	AAG Lys	ATT	AAT	. cc	: TC	c	520 1620
162 541	C)	GA A y L	AA ys		CCA Pro	ACC Thr	ACC	TTC Phe	CCC Pro	AAC Lys	GAT Asp	TAI	TC:	G GA		गः (	•	TCC	TGG		מלם	Ser CC.		540 1680
168 561	G(	SA G	AG Iu	CCA Pro	AAG Lys	GAC Asp	AAT Asn	CCC	CAA Gla	AGA	GTG	ज व	; TAC	. GA	G GA	u c	SAC	Ser	TAC	Thr GTG	Phe	Pro TAC	. ,	560 1740
						ACC.	TTC	GCT	GIG	GAA	V 41	4 #1	1 77	Glu	Ch	L) A	-	li.	Tyr	Val	Giy	. Tyr	-	580
	AC	A A.	.G 1	П	GAA	TAC	444	· ·				~4	ı yr	Clu	Pho		ily	Тут	GGC Gly	CTC Lev	TCT Ser	TAC Tyr	-	800 600
						•	_,-	. —		Lys	110	A la	lle	Απο	Giv		AG lu	Th-	CTC Lev	AGA Arg	رار درو	TCG Ser	-	860 620
								•		AGA Arg	7718	Uiy	Lys	Giu	Val	5-		71-	GTC Val	TAC Tyr	ATC	AAA Lys		920 540
								•	-,-	CCC Pro	1 116	UIR	CIB	Leu	Lvs	A 1:	. Z	N	CAC His		ACA	٨١٨	19	80
661 1981	Lev	Leu	S A.	AC C	CG ro	GGT Gly	GAA Giz	TCA Ser	GAA Glu	GAA Giu	ATC lle	TCC Ser	TTG Lev	GAA	ATT	, cc	T (	тс	AGA	Lys GAT	Dur	Lys GCG	20-	40
2041 681	AGT Ser	TTC Phe	G/	AT G	GG ,	AAA	GAA Glu	TGC (	<del>जा</del>	GTC Val	GAG -	TCA	GGA	GAA	lic TAC	Pro	.G G	ru TC	Arg AGG	Asp GTC	Leu GGT	Ala		80
2101 701					AT /		NGG ·	TG A	GA (	GAT .	ATT	361 TTT	CTG	Glu	Twe	C1			Arg	Vai	Ciy	GCA Ala	70	<b>x</b> 0
2161		TGA	216	<b>56</b>		- •	₩ \$ (	- U -	irg /	Asp	lle !	Phe	Lev	Val 	Glu	Gly	Gi			AGA Arg	TTC Phe	Lys	72	

Figure 5b(Continued)

#### THERMOCOCCUS AEDII12RA GLYCOSIDASE COMPLETE GENE BEQUENCE - 9/95 ATG ATC CAC TGC CGG GTT ANA GGG ATT ATA TCT GAG GCT CGC GGC ATA ACC ATC ACA ATA Met lie His Cys Pro Val Lys Gly lie lie Ser Glu Ala Arg Gly lie Thr lie Thr lie 60 61 GAT TTA AGT TIT CAA GGC CAA ATA AAT TTG GTG AAT GCT ATG ATT GTC TIT CCG GAG 20 Asp Leu Ser Phe Gin Cly Gin Ile Asn Asn Leu Val Asn Ala Het Ile Val Phe Pro Glu 120 121 TTC TTC CTC TTT GGA ACC GCC ACA TCT TCT CAT CAG ATC GAG GGA GAT AAT AAA TGG AAC 40 41 Phe Phe Leu Phe Gly Thr Ale Thr Ser Ser His Gln Ile Glu Gly Asp Asn Lys Trp Asn 180 CAC TGG TGG TAT TAT GAG GAG ATA GGT AAG CTC CCC TAC AAA TCC GGT AAA GCC TGC AAT 181 60 Asp Trp Trp Tyr Tyr Glu Glu Ile Gly Lys Leu Pro Tyr Lys Ser Gly Lys Ala Cys Asn 61 240 CAC TOG GAG CTT TAC AGG GAA GAT ATA GAG CTA ATG GCA CAG CTC GGC TAC AAT GCC TAC 80 His Trp Glu Leu Tyr Arg Glu Asp Ile Glu Leu Het Ala Gln Leu Gly Tyr Asn Ala Tyr 300 301 CGC TIT TCG ATA GAG TGG AGC CGT CTC TTC CCG GAA GAG GGC AAA TTC AAT GAA GAA GCC 100 101 Arg Phe Ser Ile Glu Trp Ser Arg Leu Phe Pro Glu Glu Gly Lys Phe Asn Glu Glu Ala TTC AAC CGC TAC CGT GAA ATA ATT GAA ATC CTC CTT GAG AAG GGG ATT ACT CCA AAC GTT 120 Phe Asn Arg Tyr Arg Glu Ile Ile Glu Ile Leu Leu Glu Lys Gly Ile Thr Pro Asn Val 420 ACA CTG CAC CAC TTC ACA TCA CCG CTG TCG TTC ATG CGG AAG CGA GGC TTT TTG AAG GAA 140 Thr Leu His His Phe Thr Ser Pro Leu Trp Phe Het Arg Lys Gly Gly Phe Leu Lys Glu 480 CAA AAC CTC AAG TAC TGG GAG CAG TAC GTT GAT AAA GCC GCG GAG CTC CTC AAG GGA GTC 160 Glu Asn Leu Lys Tyr Trp Glu Gln Tyr Val Asp Lys Ala Ala Glu Leu Leu Lys Gly Val 540 ANG CTT GTA GCT ACA TTC ANC GAG CCG ATG GTC TAT GTT ATG ATG GGC TAC CTC ACA GCC 180 541 Lys Leu Val Ala Thr Phe Asn Glu Pro Het Val Tyr Val Het Het Gly Tyr Leu Thr Ala 600 דאכ זכם ככם כככ דדכ אדכ אאם אכד כככ דדד אאא פכר דדד אאא פדד פככ פכא אאכ כדכ כדד 200 Tyr Trp Pro Pro Phe Ile Lys Ser Pro Phe Lys Ala Phe Lys Val Ala Ala Asn Leu Leu 201 660 ANG GCC CAT GCA ATG GCA TAT GAT ATC CTC CAT GGT AAC TIT GAT GTG GGG ATA GTT AAA 220 661 Lys Ala His Ala Het Ala Tyr Asp Ile Leu His Gly Asn Phe Asp Val Gly Ile Val Lys 221 720 721 AND ATO COO ATA ATO CTO COT GOA AGO AND AGA GAG ANA GAG GTA GAA GOT GOO CAA AAG 240 Asn Ile Pro Ile Met Leu Pro Ala Ser Asn Arg Glu Lys Asp Val Glu Ala Ala Gln Lys 780 GCG GAT AAC CTC TIT AAC TGG AAC TTC CTT GAT GCA ATA TGG AGC GGA AAA TAT AAA GGA 260 781 Ala Asp Asn Leu Phe Asn Trp Asn Phe Leu Asp Ala Ile Trp Ser Gly Lys Tyr Lys Gly 280 GCT TIT GGA ACT THE ANA ACT CON GAN AGE GAT GCA GAC TTC ATA GGG ATA AND THE THE 841 Ala Phe Gly Thr Tyr Lys Thr Pro Glu Ser Asp Ala Asp Phe Ile Gly Ile Asn Tyr Tyr 281 900 300 ACA GCC AGC GAG GTA AGG CAT AGC TGG AAT CCG CTA AAG TIT TTC TTC GAT GCC AAG CTT Thr Ala Ser Glu Val Arg His Ser Trp Asn Pro Leu Lys Phe Phe Phe Asp Ala Lys Leu 960 GCA GAC TTA AGC GAG AGA AAA ACA GAT ATG GGT TGG AGT GTC TAT CCA AAG GGC ATA TAC 320 Ala Asp Leu Ser Glu Arg Lys Thr Asp Het Gly Trp Ser Val Tyr Pro Lys Gly Ile Tyr 1020 GAA GCT ATA GCA AAG GTT TCA CAC TAC GGA AAG CCA ATG TAC ATC ACG GAA AAC GGG ATA Glu Ala Ile Ala Lys Val Ser His Tyr Gly Lys Pro Het Tyr Ile Thr Glu Asn Gly Ile 1080 1081 GCT ACC TTA GAC GAT GAG TGG AGG ATA GAG TTT ATC ATC CAG CAC CTC CAG TAC GTT CAC 360 361 Ala Thr Leu Asp Asp Glu Trp Arg Ile Glu Phe Ile Ile Gln His Leu Gln Tyr Val His 1140 ANA GCC TTA ANC GAT GGC TTT GAC TTG AGA GGC TAC TTC TAT TGG TCT TTT ATG GAT AAC 380 381 Lys Ala Leu Asn Asp Gly Phe Asp Leu Arg Gly Tyr Phe Tyr Trp Ser Phe Het Asp Asn 1200 400 TTC GAG TOG GCT GAG GCT TTT AGA CCA CGC TTT GGG CTG GTC GAG GTG GAC TAC ACG ACC Phe Glu Trp Ala Glu Gly Phe Arg Pro Arg Phe Gly Leu Val Glu Val Asp Tyr Thr Thr TTC AAG AGG AGA CCG AGA AAG AGT GCT TAC ATA TAT GGA GAA ATT GCA AGG GAA AAG AAA Phe Lys Arg Arg Pro Arg Lys Ser Ala Tyr Ile Tyr Gly Glu Ile Ala Arg Glu Lys Lys 421 1120 ATA AAA GAC GAA CTG CTG GCA AAG TAT GGG CTT CCG GAG CTA TGA 440 1321 Ile Lys Asp Clu Leu Leu Ala Lys Tyr Cly Leu Pro Clu Leu End

Figure 6

# THERMOCOCCUS CHITONOPHAGUS GLYCOSIDASE - 22G COMPLETE SEQUENCE - 9/95

COMPALIE SEQUENCE - 9/95
1 TTG CTT CCA GAG AAC TIT CTC TGG GGA GTT TCA CAG TCC GGA TTC CAG TTT GAA ATG GGG 60
1 Het Leu Pro Glu Asn Phe Leu Trp Gly val Ser Gln Ser Gly Phe Glu Phe Glu Het Gly 20
and led it of the Cly Val Ser Gly Phe Gly Phe Gly Val GCC 60
61 GAC AGA CTG AGG AGG CAC ATT GAT CCA AAC ACA GAT TGG TGG TAC TGG GTA AGA GAT GAA 120
21 ASP Arg Leu Arg Arg Min Til GAT CCA AAC ACA GAT TGG TGG TAC TGG CTA ACC ACC
21 ASP ANG LEU ANG ANG HIS ILE ASP PRO ASR THE ASP TEP TEP TYE TEP VAL ANG ASP CIU 40
121 TAT AAT ATC AAA AAA GGA CTA GTA AGT GCC GAT CTT CCC GAA GAC GCT ATA AAT TCA TAT 180
41 Tyr Asn Ile Lys Lys Gly Leu Vai Ser Gly Asp Leu Pro Glu Asp Gly Ile Asn Ser Tyr 60
by Gly Leu Val Ser Gly Asp Leu Pro Gly Are GOT ATA AAT TCA TAT 180
181 GAA TT1 TIM TIM TO THE AST SET TUT AS
61 Glu Leu Tyr Glu Arg Asp Gln Glu Ile Ala Lys Asp Leu Gly Leu Asn Thr Tyr Arg Ile 80
old Arg Asp Gin Glu Ile Ala Lys Asp Lou Glu AAC ACA TAT AGG ATC 240
491 GGA ATT CAN man
81 GIV TIE CIV. TO AGE AGE TO THE COA TOG CCA ACC 100
241 GGA ATT GAA TGG AGC AGA GTA TIT CCA TGG CCA ACG ACT TIT CTC GAC GTG GAG TAT GAA 300 81 Gly Ile Glu Trp Ser Arg Val Phe Pro Trp Pro Thr Thr Phe Val Asp Val Glu Tyr Glu 100 301 ATT GAT GAG TCT TAC GCG TTC CTC ACG TTC GAC GTG GAG TAT GAA 300
301 ATT COM COLO TAT GAN 300
101 ATT GAT GAG TCT TAC GGG TTG GTA AAG GAT GTG ANG AND
301 ATT GAT GAG TCT TAC GGG TTG GTA AAG GAT GTG AAG ATT TCT AAA GAC GCA TTA GAA AAA 360 361 CTT GAT GAA ATC GCT AAG GAA AC GCA AC AC GCA AC AC GCA AC GCA AC GCA AC AC GCA AC GCA AC AC GCA AC AC GCA AC AC AC AC GCA AC A
161 cm and the Lys Asp Val Lys Ile Ser Lys Asp Ala Leu Clu Lys
161 CTT GAT GAA ATC GCT AAC CAA AGG GAA ATA ATA TAT TAT AGG AAC CTA ATA AAT TCC CTA 420
121 Leu ASP Glu Ile Ala ASP Glu Ile Ile Ile Tyr Tyr Arg ASP Leu Ile ASP Ser Leu 140
ATT ATT GIR HE THE TYT TYT ATT ASN LEW ILE ATT THE CTA 420
421 AGA AAG AGG CGT TTT AAG GTA ATA GTA ATA
421 AGA AAG AGG CGT TTT AAG GTA ATA CTA AAC CTA AAT CAT TTT ACC CTC CCA ATA TGG CTT 480  481 CAT GAT CCT ATC GAA TCT ACC CTC AAA TGG CTT 480
and the Leu Asn Leu Asn His Phe Thr Leu Pro XIA TGG CTT 480
481 CAT GAT CCT ATC GAA TCT AGA GAA AAA GCC CTG ACC AAT AAG AGA AAC GGA TGG GTA AGC 540
161 His Asp Pro Ile Glu Ser Arg Glu Lys Ala Leu Thr Asn Lys Arg Asn Cly Trp Vel Ser 180
Ser Arg Glu Lys Ala Leu Thr Asn Lys Arg asn Clu To GTA AGC 540
541 GAN AGG AGT GTT ATA GAG TIT GCA ANA TIT GCC GCG TAT TTA GCA TAT ANA TIT GGA GAC 600
181 Glu Arg Ser Val Tie Clu at GCA AAA TIT GCC GCG TAT TTA GCA TAT AND
and the Ala Lys Phe Ala Ala Tyr Leu Ala Ty
181 Glu Arg Ser Val Ile Glu Phe Ala Lys Phe Ala Ala Tyr Leu Ala Tyr Lys Phe Gly Asp 200 601 ATA GTA GAC ATG TGG AGC ACA TTT AAT GAA CCT ATG GTG GTC GCC GAG TTG GGG TAT TTA 660 201 Ile Val Asp Het Trp Ser Thr Phe Ash Glu Pro Het Val Val Ala Glu Ing 660
201 Ile Val Asp Het Trp Ser Thr Phe Ash Glu Pro Het Val Val Ala Glu Leu Gly Tyr Leu 220
Ser Thr Phe Asn Glu Pro Het Val Val Ala Clu TTG GGG TAT TTA 660
661 GCC CCA TAC TO 220
661 GCC CCA TAC TCA GCA TTC CCC CCG GCA GTC ATG AAT CCA GAA GCA GCA AAG TTA GTT ATG 221 Ala Pro Tyr Ser Gly Phe Pro Pro Gly Val Het Asn Pro Glu Ala Ala Lys Leu Val Het 240  721 CTA CAT ATG ATA AAG CCC CAR CCC CCG GCA GTC ANG TTA GTT ATG 720
720 THE PRO PRO Gly Val Met Ash Pro Gly Ala ALA AAG TTA GTT ATG 720
721 CTA CAT ATC ATT ATC ATC
241 Leu His Het Ile Asn Ala His Ala Leu Ala Tyr Arg Het Ile Lys Lys Phe Asp Arg Lys 260
ASA ALA HIS Ala Leu Ala Tyr Ary Met Ile Luc Tir GAC AGA AAA 780
781 AAA GCT GAT COL SAN COL SAN COL SAN ASP AFG Lys 260
781 AAA GCT GAT CCA GAA TCA AAA GAA CCA GCT GAA ATA GGA ATT ATA TAC AAT AAC ATC GGC 840 261 Lys Ala Asp Pro Glu Ser Lys Glu Pro Ala Glu Ile Gly Ile Ile Tyr Asn Asn Ile Gly 280 841 GTC ACA TAT CCG TTT AND GOOD
FIG. Ser Lys Glu Pro Ala Glu Ile Gly Ile The TAT AAC ATC GGC 840
841 GTC ACA TAT GGG
841 GTC ACA TAT CCG TTT AAT CCG AAA GAC TCA AAG GAT CTA CAA GCA TCC GAT AAT GCC AAT 900 281 Val Thr Tyr Pro Phe Asn Pro Lys Asp Ser Lys Asp Leu Gln Ala Ser Asp Asn Ala Asn 300 901 TTC TTC CAC AGT GCC GTA TTC
FIG Fire Ash Pro Lys Asp Ser Lys Asp Leu Cla Ala CC GAT AAT GCC AAT 900
901 TTC TTC CAS AST Ala AST 300
301 Phe Phe His Con GGG CTA TTC TTA ACG GCT ATC CAC AGG GCA AND
901 TTC TTC CAC AGT GGG CTA TTC TTA ACG GCT ATC CAC AGG GGA AAA TTA AAT ATC GAA TTT 960 961 GAC GGA GAG ACA TTT CTT TA ACG GCT ATC CAC AGG GGA AAA TTA AAT ATC GAA TTT 960
961 GAC GGA CAC ACT TO THE STATE OF THE STAT
321 ASD GIV GIV GIV TAC CIT TAC TAT TTA ANG GGG AND GIVE
961 GAC GGA GAG ACA TIT GIT TAC CIT CCA TAT TIA AAG GGC AAT GAT TGG CTG GGA GTG AAT 1020  1021 TAT TAT ACA AGA GAA GTG CTG GGA GTG ASD 340
1021 TAT TAT ACT ACT ACT ACT ACT ACT ACT ACT
1021 TAT TAT ACA AGA GAA GTC GTT AAA TAC CAA GAT CCC ATG TTT CCA AGT ATC CCT CTC ATA 1080
THE THE ARG GIU VAI VAI LYS TYP GID ASD PED MAN THE COA AGE ATC CCT CTC ATA 1080
341 Tyr Tyr Thr Arg Glu Val Val Lys Tyr Gln Asp Pro Het Phe Pro Ser Ile Pro Leu Ile 360
1081 AGC TTC AAG GGC GTT CCA GAT TAT GGA TAC GGA TGT AGA CCA GGA ACG ACG TCA AAG GAC 1140  1081 AGC TTC AAG GGC GTT CCA GAT TAT GGA TAC GGA TGT AGA CCA GGA ACG ACG TCA AAG GAC 1140
361 Ser Phe Lys Gly Val Pro Asp Tyr Gly Tyr Gly Cys Arg Pro Gly Thr Thr Ser Lys Asp 380
1141 GGT AAT CUT COT ASS 380
1141 GGT AAT CCT GTT AGT GAC ATT GGA TGG GAG GTA TAT CCC AAA GGC ATG TAC GAC TCT ATA 1200
401 Val Ala Ala Asn Glu Tyr Gly Val Bro
1261 AAA GAT GTA TTA AGG CCC TAT TAC ATC GCA TCT CAC ATT GAA GCC ATG GAA GAG GCT TAC 1320 421 Lys Asp Val Leu Arg Pro Tyr Tyr Ile Ala Ser His Ile Glu Ala Met Glu Glu Ala Tyr 440
FIG TYP TYP IIe Ala Ser His IIe Glu Ala Het Glu Glu Ala The
net did did Ala Tyr 440

Figure 7a

1321	CLU ABN GI	TAT C	AC CTC		·													
441	Clu Asn Gi	TVE	ASD VAL	767	GGA	TAC	TTA	CAC	TCC	GCA	TTA	ACC	GAT	AAT	TAC	C: 8.8	700	
						-	_			~	rev	Thr	Asn	Acn	TV	C1		Linn
																		460
461	Ala Leu Gly	Phe A	IC Met	A	117	CCC	TTC	TAC	CAA	CTA	AAC	TTG	ATA	ACC	AAA	GAG	101	
						•	-				~*11	ren	110	The	1	C1		1440
	TOOL CLE NOW		· ·															480
481	Lys Pro Arg	Lys L	VS Ser	V-1	100 I	CTA	TTC	AGA	CYC	ATA	CTT	ATT	AAT	AAT	GCG	CTA	ACA	1500
							-				Val	Ile	Asn	Asn	Glv	Leu	The	
	ACC AVC YES	ACC A															••••	500
501	Ser Asn Ile	Arg L	vs Glu	Tla 1		DAG	CAC	CCC	TAG	15	36							
				***	reg (	-10	CIU	GIA	End	51	2							

Figure 7b(Continued)

# PYROCOCCUS FURIOSUS GLYCOSIDASE - 701 - COMPLETE GENE SEQUENCE - 10/95

1 ATC TTC COMPLETE CHIEF SEQUENCE - 10/95	
1 ATG TTC CCT GAA AAG TTC CTT TGG GGT GTG GCA CAA TCG GGT TTT CAG TTT GAA ATG GGG 61 GAT AAA GTG ATG	
Lys Phe Leu Trp Gly Val Ala Gly see the TIT CAG TTT GAA ATG CCC	
1 Het Phe Pro Glu Lys Phe Leu Trp Gly Val Ala Gln Ser Gly Phe Gln Phe Glu Het Gly	60
22 Asp Lys Leu Arg Art ATT GAC ACT AAC ACT GAT TOO	50
61 GAT AAA CTC AGG AGG AAT ATT GAC ACT AAC ACT GAT TGG TGG CAC TGG GTA AGG GAT AAG 22 Asp Lys Leu Arg Arg Asn Ile Asp Thr Asn Thr Asp Trp Trp His Trp Val Arg Asp Lys 121 ACA AAT ATA GAG AAA GGC CTC GTT AGT GGA CAT GGT CTC GTT AGT TGA CAT TAR AND TAR THE ASP TAR CAT GGA	
	120
41 The Ash Ile Gly Law GC CTC GTT AGT GGA GAT CTT CCC GIA	40
121 ACA ART ATA GAG ARA GGC CTC GTT AGT GGA GAT CTT CCC GAG GAG GGG ATT AAC AAT TAC  41 Thr Ash Ile Glu Lys Gly Leu Val Ser Giy Asp Leu Pro Glu Glu Gly Ile Ash Ash TAC  181 GAG CTT TAT GAG ARAG GAC CAT GAG ATT GCA GAG ATT GGA GAG GAG GAG GAG ATT AAC AAT TAC  61 Gly Leu Tath GAG ARAG GAC CAT GAG ATT GCA GAG ATT GCA GAG ATT GCA GAG ATT GCA GAG GAG GAG GAG GAG GAG GAG GAG GAG	
181 GAG CTT TAT GAG AAG GAG GAG GAG GAG GAG GAG G	100 60
181 GAG CTT TAT GAG AAG GAC CAT GAG ATT GCA AGA AAG CTG GGT CTT AAT GCT TAC AGA ATA 61 Glu Leu Tyr Glu Lys Asp His Glu Ile Ala Arg Lys Leu Gly Leu Asn Ala TAC 241 GGC ATA GAG TAC	50
The state of the s	240
61 Glu Leu Tyr Glu Lys Asp His Glu Ile Ala Arg Lys Leu Gly Leu Asn Ala Tyr 241 GGC ATA GAG TGG AGA ATA TTC GCA TGG TGG TGG AGA ATA TTC GCA TGG TGG TGG AGA ATA TTC GCA TGG TGG TGG TGG TGG TGG AGA ATA TTC GCA TGG TGG TGG TGG TGG TGG TGG TGG TGG TG	80
241 GGC ATA GAG TGG AGC AGA ATA TTC GCA TGG CCA ACG ACA TTT ATT GAT GTT GAT TAT AGC 301 TAT ATT GAT GAT GAT GAT TAT AGC	
301 THE PEO TEP PEO The The Phe Ile and GIT GAT TAT AGC	300
81 Gly Ile Glu Trp Ser Arg Ile Phe Pro Trp Pro Thr Thr Phe Ile Asp Val Asp Tyr Ser 101 Tyr Asn Glu Ser Tyr Asn Leu Ile Glu Asp Val Lys Ile Thr Lys Asn Dr Lyr Asn Cas GAG GAG 361 TTA Car Cas	100
101 Tyr Asn Glu Ser Tyr Asn Leu Ile Glu Asp Val Lys Ile Thr Lys Asp Thr Leu Glu Glu Il	
361 The Car and The Gas Gas 3	160
121 Lau An GAG ATC GCC AAC AAG AGG GAG CTC CTC CCC	20
361 TTA GAT GAG ATC GCC AAC AAG AGG GAG GTC GCC TAC TAT AGG TCA GTC ATA AAC AGC CTG 4 121 Leu Asp Glu Ile Ala Asn Lys Arg Glu Val Ala Tyr Tyr Arg Ser Val Ile Asn Ser Leu 1	
421 AGE AGE AND AGE CTG 4	20
421 AGG AGG AAG GGG TTT AAG GTT ATA GTT AAT CTA AAT CAC TTC ACC CTT CCA TAT TGG TTG 40 CAT CAT CAT CAT CAT CAT CAT CAT CAT TTC ACC CTT CCA TAT TGG TTG 40 CAT	40
The Lys Val 11e Val 11m AAT CAC TTC ACC CTT CCA TAT TCC TTC	_
	90
161 His Asp Pro 11 cl SCT ASG GAG AGG GCG TTA ACT 120	60
48: CAT GAT CCC ATT GAS GCT AGG GAG AGG GCG TTA ACT AAT AAG AGG AAC GGC TGG GTT AAC 54 161 Kis Asp Pro 11e Glu Ala Arg Glu Arg Ala Leu Thr Ash Lys Arg Ash Gly Trp Val Ash 161 542 CCA AGA ACA GTT ATA GAG TTT GGA AAG TAT GGG AAG TAT GGA	
54: CCA AGA ACA GTT BTD COM	
54: CCA AGA ACA GTT ATA GAG TIT GGA AAG TAT GGC GGT TAC ATA GGC TAT AAG TIT GGA GAT 181 191 Pro Arg Thr Val Ile Glu Phe Ala Lys Tyr Ala Ala Tyr Ile Ala Tyr Lys Phe Gly Asp 20 201 Ile Val Arg TGG AGC AGG TIT AAT GAG GGT TAC ATA GCC TAT AAG TIT GGA GAT 60	. 0
SCI ITS THE BLU Phe Ala Lys Tyr Ala Ala Tyr Ila Ala TYR GCC TAT AMG TIT GGA GAT 60	10
SOI ATA GTG GAT ATG TGG AGC ACC TTT AND THE ATT THE ALE TYP IIE ALE TYP Lys Phe Gly Asp 20	
201 Ile Val Asp Met TID Ser The GAG CCT ATG GTG GTT GTT COO	•
601 ATA GTG GAT ATG TGG AGC ACG TIT AAT GAG CCT ATG GTG GTT GTT GAG CTT GGC TAC CTA 66 201 Ile Val Asp Met Trp Ser Thr Phe Ash Glu Pro Met Val Val Glu Leu Gly Tyr Leu 22 221 A's Pro TAC TCT GGC TTC CCT CCA GGG GTT GTT CTT CAG GTU Leu Gly Tyr Leu 22	0
721 NOT THE TET GGC TTC CCT CC3 CCC CC3	٥
661 GCC CCC TAC TCT GGC TTC CCT CCA GGG GTT CTA AAT CCA GAG GCC GCA AAG CTG GCG ATA  721 CTT CAC ATG ATA AAT GCA CAT GCT TTA CCT TAC ATG ATG ATA AATA A	
721 CTT CAC AND A COLOR OF THE STATE OF THE	
241 Lau Nie ATA AAT GCA CAT GCT TTA GCT TITA GCT	٥
721 CTT CAC ATG ATA AAT GCA CAT GCT TTA GCT TAT AGG CAG ATA AAG AAG TTT GAC ACT GAG 241 Leu His Her Ile Asn Ala His Ala Leu Ala Tyr Arg Gln Ile Lys Lys Phe Asp Thr Glu 261 LAA GCT GAT AAG GAT TCT AAA GAG CCT CCA CAS ACT GAG ACT GAG ACT GAG ACT GAT AAG GAT AAG GAT TCT AAA GAG CCT CCA CAS ACT GAT AAG GAT TCT AAA GAG CCT CCA CAS ACT GAT AAG GAT TCT AAA GAG CCT CCA CAS ACT GAT AAG GAT TCT AAA GAG CCT CCA CAS ACT GAT AAG GAT TCT AAA GAG CCT CCA CAS ACT GAT AAG GAT TCT AAA GAG CCT CCA CAS ACT GAT AAG GAT TCT AAA GAG CCT CCA CAS ACT GAT AAG AAG ACT GAT AAG GAT TCT AAA GAG CCT CCA CAS ACT GAT AAG GAT TCT AAA GAG CCT CCA CAS ACT GAT AAG AAG AAG ACT GAT AAG GAT TCT AAA GAG CCT CCA CAS ACT GAT AAG AAG AAG ACT GAT AAG AAG AAG AAG AAG AAG AAG AAG AA	_
261 Lys Ala Ala TAN THE STATE OF THE STATE O	,
781 AAA GCT GAT AAG GAT TCT AAA GAG CCT GCA GAA GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT GGT GGT ATA CCC AAG GAT CCG AAC GTT TGG 110 GTT GGT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT GGT GGT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT GGT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT GGT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC ATT GGA 840 GTT GGT ATA ATT TAC AAC ATT GGT ATA ATT TAC AAC ATT GGT ATT TAC AAC ATT TAC AAC ATT TAC AAC ATT TAC AAC A	,
841 GTT GCT TAT CCC AAG GAT CCG AAC GAT TCC AAG GAT GTT AAG GCA GCA GAA AAC GAC AAC 900	
261 Val Ala Tyr Pro Lys Asp Pro Asn Asp Ser Lys Asp Val Lys Ala Ala Glu Asn Asp Asn 300  901 TTC TTC CAC TCA GGG CTG TTC TTC GAG CCC AST CTC AND GAL Ala Glu Asn Asp Asn 300	
ASD ASD ASD Ser Lys Asp Val Lys Ala GAR	
901 TTC TTC CAC TCA GGG CTG TTC TTC GAG GCC ATA CAC ANA GGA ANA CTT ANT ATA GAG TTT 960 961 GAC GGT GAA ACG TTT ATA GAT GCC CCC TTT GTT ATA GAG TTT 960 951 AND CLU GAA ACG TTT ATA GAT GCC CCC TTT GTT ATA GAG TTT 960 951 AND CLU GAA ACG TTT ATA GAT GCC CCC TTT GTT ATA GAT GCC ATA GTT ATA GAT GCC GCC TTT GTT ATA GAT GCC CCC TTT GTT ATA GAT GCC GCC GCC GCC GCC GCC GCC GCC GCC GCC GC	
961 GRC COT COL	
The Phe Ile Asp Ala Pro TAT CTA AAG GGC AAT GAC TGG ATA GGC	_
1021 TAC TAC ACA AGG GAA GTA GTT ACG TAT CAG GAA CCA ATG TTT CCT TCA ATC CCG CTG ATC  1081 ACC TTT ACG GAU Val Val Thr Tyr Gln Glu Pro Met Phe Pro Ser Ile Pro LTG ATC  1080 ACC TTT ACG GAU CTA GTA GTT ACG GAU CCA ATG TTT CCT TCA ATC CCG CTG ATC  1081 ACC TTT ACG GAU Val Val Thr Tyr Gln Glu Pro Met Phe Pro Ser Ile Pro LTG ATC  1080 ACC TTT ACG GAU CTA GTT ACG GAU CTA GAU CCA GTT CCA ATC CCG CTG ATC  1080 ACC TTT ACG GAU CTA GTT ACG GAU CTA GTT ACG GAU CTA GAU	
341 Tyr Tyr Thr Arg Glu Val Val Thr Tyr Gln Glu Pro Met Phe Pro Ser Ile Pro Leu Ile 360	
	,
1081 ACC TTT AAG GGA GTT CAA GGA TAT GGC TAT GCC TGC AGA CCT GGA ACT CTG TCA AAG GAT 1140  1141 GAC AGA CCC GTC AGC GAC ATA GGA TGG CAA CCT GGA ACT CTG TCA AAG GAT 1140  1141 GAC AGA CCC GTC AGC GAC ATA GGA TGG CAA CCT GGA ACT CTG TCA AAG GAT 1140  381 AAD ACC GTC AGC GAC ATA GGA TGG CAA CCT GCA ACT CTG TCA AAG GAT 380	
THE GIR GIY TYP GIY TYP ALA CYS AND PRO GON ACT CTG TCA AAG CAT 1140	,
381 Asp Arg Pro Val Ser Arm 110 CLA CAG GAG GAG GAG GAG GAG GAG GAG GAG GA	
1141 GAC AGA CCC GTC AGC GAC ATA GGA TGG GAA CTC TAT CCA GAG GGG ATG TAC GAT TCA ATA 1200 1201 GTT CAB CCT CAG CCT CAG GAC ATA GGA TGG GAA CTC TAT CCA GAG GGG ATG TAC GAT TCA ATA 1200	
401 Val Giu Ala His Lys Tyr Gly Val Pro Val Tyr Val Thr Glu Ash Gly He Ala Ash Ser 420	
The Glu Ash Glu Ash Glu Ash Glu Ash Glu Ash Glu GAT TCA 1260	
420 Ala Ang Ser 420	

Figure 8a

	Lys Asp					-						116	Lyg	мес	110	Glin	1	A : -	nh -	1320
	Glu Asp	Gly	Tyr	eyr cyy	GII Val	Lys	GGC.	TAC Tyr	TTC Phe	K73 CYC	TGG Trp	GCA Ala	TTA Leu	ACT	GAC	AAC	TTC	GAG	TGG	1380 160
461	Ala Leu	CIY	Phe	AGA Arg	ATG Me t	<b>CGC</b>	TTT Phe	CCC Cly	CTC Leu	TAC Tyr	GAX Glu	STC Val	AAC Asn	CTA Leu	ATT Ile	ACA	AAG	GAG	AGA	1440
481	Ile Pro	Arg	C) r CYC	AAG Lys	<b>Set</b>	GTG Val	TCG Ser	ATA Ile	TTC Phe	AGA Arg										1500
1201	Lys Lys	AIT	GAA	CAC	C 2 2				_			33					•			230

Figure 8b(Continued)

#### Bankia gouldi endoglucanase (37071)

(37091)
9 18 27 36 AS
5' ATG AGA ATA COT TTA COS 300 45 54
Met Arg Ile Arg Leu Ala The Cit GCG CTC TGC GCA GCG CTG AGC CCA GTC AGC
Met Arg Ile Arg Leu Ala Thr Leu Ala Leu Cys Ala Ala Leu Ser Pro Val Thr
0.5
TTT CCA GAT AAT CTA ACC
Phe Ala ASD ASD VALUE GTA CAA ATC GAC GCC GAC GGC GGT AND AND
Phe Ala Asp Asn Val Thr Val Glm Ile Asp Ala Asp Gly Cly Lys Lys Leu Ile
AGC CGA GCC CTT TAC GGC ATG AAT AAC TCC AAC GCA CAA AGC CTT ACC GAT ACT Ser Arg Ala Leu Tyr Gly Met Asn Asn Ser Asn Ala Gly Ser Act
Ser Arg Ala Louis GGC ATG ART AAC TCC AAC GCA GAA AGC CTM 162
Ser Arg Ala Leu Tyr Gly Met Asn Asn Ser Asn Ala Glu Ser Leu Thr Asp Thr
171 180
GAC TYPE CAP COM 189 198 200
CAC TGG CAG CGT TIT CGC GAT GCA GGT GTG CGC ATG CTG CGG GAA AAT GGC GGC ABP Trp Gln Arg Phe Arg Asp Ala Gly Val Arg Met Lev Arg CTG CGG GAA AAT GGC GGC
And the Arg Phe Arg Asp Ala Gly Val Arg Mor Louis Cod GAA AAT GGC GGC
and Arg Glu Asn Gly Glv
443 774
701 ACC ACC 121 mm 22 261 270
ABN ABN SET THE LYS TYP ASH TED GIR LET HIS AND AGT CAT CCG GAT TGG
and her set will blo yeb ind
4/3 200
AND AND ANT GTC TAC COO CO. 315
TAC AAC AAT GTC TAC GCC GGC AAC AAC AAC TGG GAC AAC CGG GTA GCC CTG ATT  TYT ASD ASD Val TYT Ala Gly Asd Asd Asd TSD Asd Asd Acc CGG GTA GCC CTG ATT
The same will be seen the same same same same same same same sam
CAG GAA AAC CTC CCC CCC CCC
CAG GAA AAC CTG CCC GGC GCC GAC ACC ATG TGC GCA TTC CAG CTC ATC GGT AAG Glm Glm Asm Leu Pro Gly Ala Asp Thr Net TED Ala Pho Gla CTC ATC GGT AAG
Gln Glu Asn Leu Pro Gly Ala Asp Thr Mer Trp Ala Phe Gln Leu Ile Gly Lys
28/ 38c
OLU GCG ACT TOTAL COLUMN COLUM
VAL Ala The See Ala TVE As Bhalant TGG GAA TTC AAC CAG TCG CAA
The Ash Gln Ser Gln
100 ACC CCC CCC CCC CCC A77
TEP TEP The Gly Val Ala Gle Asp Lot all GCC GCC GCT GAA CCC AAT CTG GAC
Trp Trp Thr Gly Val Ala Gln Asn Leu Ala Gly Gly Gly Glu Pro Asn Leu Asp
N23 FA/
GGC GGC GAA COT COT 523 522 531 532
Gly Gly Glu Ala Leu Val Glu Gly Asp Pro Ash Leu Tyr Leu Het Asp Trp
off the App Pro Ash Leu Tyr Leu Met Ash Tro
TES CEA GCC GAC ACM CON THE STATE SAE
Ser Pro Ala Asp Thr Val Gly Tla CRC CAC TGG TTT GGC GTA AAC GGG CTG
Ser Pro Ala Asp Thr Val Gly He Leu Asp His Trp Phe Gly Val Asn Gly Leu
943 215
CCC GTG CGG CGT CCC 111 CCC
Gly Val Arg Arg Gly Lyr Ala TAC TGG AGT ATG GAT AAC GAG CCC CGC ATG
Gly Val Arg Arg Gly Lys Ala Lys Tyr Trp Ser Met Asp Asn Glu Pro Gly Ile
TGG GTT GGC ACC GGG 675 684 683
TCG GTT GGC ACC CAC GAC GAT GTA GTG AAA GAA CAA ACG CCG GTA GAA GAT TTC  TTP Val Gly Thr His Asp Asp Val Val Lye Gly Gly Thr Brown GAA GAT TTC
Trp Val Gly Thr His Asp Asp Val Val Lys Glu Gln Thr Pro Val Glu Asp Phe
TTO VAL GIU ASD Phe
TO 2

Figure 9a

# Bankia gouldi endoglucanese (37071) (continued)

720 CTG CAC ACC TAT TTC GAA ACC GCC AAA AAA GCC CGC GCC AAA TTT CCC GGT ATT Leu His Thr Tyr Phe Glu Thr Ala Lys Lys Ala Arg Ala Lys Phe Pro Gly Ile 783 -ANA ATC ACC GGT CCG GTG CCC GCT AAT GAG TGG CAG TGG TAT GCC TGG GGC GGT Lys Ile Thr Gly Pro Val Pro Ala Asn Glu Trp Gln Trp Tyr Ala Trp Gly Gly 828 TTC TCG GTA CCC CAG GAA CAA GGG TTT ATG AGC TGG ATG GAG TAT TTC ATC AAG 837 Phe Ser Val Pro Gln Glu Gln Gly Phe Met Ser Trp Met Glu Tyr Phe Ile Lyr 882 CGG GTG TCT GAA GAG CAA CGC GCA AGT GGT GTT CGC CTC CTC GAT GTA CTC GAT 891 Arg Val Ser Glu Glu Gln Arg Ala Ser Gly Val Arg Leu Asp Val Leu Asp 927 936 CTG CAC TAC TAC CCC GGC GCT TAC AAT GCG GAA GAT ATC GTG CAA TTA CAT CGC Leu His Tyr Tyr Pro Gly Ala Tyr Asn Ala Glu Asp Ile Val Gln Leu His Arg 990 ACG THE THE GAC COC GAC THE STE TEA CHE GAT GOO AAC GGG GTG AAA ATG GTA 999 Thr Phe Phe Asp Arg Asp Phe Val Ser Leu Asp Ala Asn Gly Val Lym Met Val 1035 GAA GGT GGC TGG GAT GAC AGC ATC AAC AAG GAA TAT ATT TTC GGG CGA GTG AAC 1053 Glu Gly Gly Trp Asp Asp Ser Ile Asn Lys Glu Tyr Ile Phe Gly Arg Val Asn 1098 1107 GAT TOG CTC GAG GAA TAT ATG GGG CCA GAC CAT GGT GTA ACC CTG GGC TTA ACC Asp Trp Leu Glu Glu Tyr Met Gly Pro Asp His Gly Val Thr Leu Gly Leu Thr 1152 GAA ATG TGC GTG CGC AAT GTG AAT CCG ATG ACT ACC GCC ATC TGG TAT GCC TCC 1161 Glu Met Cys Val Arg Asn Val Arn Pro Met Thr Thr Ala Ile Trp Tyr Ala Ser 1206 1215 ATG CTC GGC ACC TTC GCG GAT AAC GGC GTC GAA ATA TTC ACC CCA TGG TGC TGG Met Leu Gly Thr Phe Ala Asp Asn Gly Val Glu Ile Phe Thr Pro Trp Cys Trp 1251 1260 1269 AAC ACC GGA ATG TGG GAA ACA CTC CAC CTC TTC AGC CGC TAC AAA CCT TAT Asn Thr Gly Met Trp Glu Thr Leu His Leu Phe Ser Arg Tyr Asn Lys Pro Tyr 1314 1323 CGG GTC GCC TCC AGC TCC AGT CTT GAA GAG TTT GTC AGC GCC TAC AGC TCC ATT Ary Val Ala Ser Ser Ser Ser Leu Glu Glu Phe Val Ser Ala Tyr Ser Ser Ile ANC GAN GCA GAN GAC GCC ATG ACG GTA CTT CTG GTG AAT CGT TCC ACT AGC GAC 1377 Asn Glu Ala Glu Asp Ala Met Thr Val Leu Leu Val Asn Arg Ser Thr Ser Glu

Figure 9b(Continued)

## Bankia gouldi endoglucanase (370P1) (continued)

1413 1422 1431 1440 1449 1458
ACC CAC ACC GCC ACT GTC GCT ATC GAC GAT TTC CCA CTG GAT GGC CCC TAC CGC
Thr His Thr Ala Thr Val Ala Ile Asp Asp Phe Pro Lou Asp Gly Pro Tyr Arg

1467 1476 1485 1494 1503 1512
ACC CTG CGC TTA CAC AAC CTG CCG GGG GAG GAA ACC TTC GTA TCT CAC CGA GAC
Thr Leu Arg Leu His Asn Leu Pro Gly Glu Glu Thr Phe Val Ser His Arg Asp

1521 1530 1539 1548 1557 1566
AAC GCC CTG GAA AAA GGT ACA GTG CGC GCC AGC GAC AAT ACG GTA ACA CTG GAG
AER Ala Leu Glu Lys Gly Thr Val Arg Ala Ser Aep Aer Thr Val Thr Leu Glu

1575 1584 1593 1602 1611
TTG CCC CCT CTG TCC GTT ACT GCA ATA TTG CTC AAG GCC CGG CCC TAA 3'
Leu Pro Pro Leu Ser Val Thr Ala Ile Leu Leu Lys Ala Arg Pro ***

Pigure 94 (Continued)

## Thermotoga maritima Alpha-oalactosidade Complete Gene Sequence (1 c + 3)

· · ·
5' GTC ATC TOT GTG GAA ATA TITC GGA AAG ACC TTC AGA GAA GGA AGA TTC GTT CTC
The Cys val Glu Ile Phe Gly Lys The Phe Arg Glu Gly Arg Phe Val Leu
ANA GAG ANA AND THE ACA CIT GAG THE GAG GAG ANG ATA CAC CITY COM
Dys Ach Phe Thr Val Glu Phe Ala Val Glu Lys Ile His Leu Gly Trp
ANG ATC TCC GGC AGG GTG ANG GGA AGT CCG GGA AGG CTT GAG GTT CTT CGA AGG
Lys Ile Ser Gly Arg Val Lys Gly Ser Pro Gly Arg Leu Glu Val Leu Arg Thr
ANA GCA CCG GNA ANG GTA CTT GTG ANC ANC TOG CAG TOC TOG GGA CCG TGC AGG Lys Ala Pro Glu Lys Val Leu Val Asn Asn Trp Gln Ser Trp Gly Pro Cys Arg
GTG GTC GAT GCC TIT TCT TTC AAA CCA CCT GAA ATA GAT CCG AAC TGG AGA TAC
val Val Asp Ala Phe Ser Phe Lys Pro Pro Glu Ile Asp Pro Am Trp Ary Tyr
ACC GCT TCG GTG GTG CCC GAT GTA CTT GAA AGG AAC CTC CAG AGC GAC TAT TTC
The Ala Ser Val Val Pro Asp Val Lou Glu Ary Asm Leu Glm Ser Asp Tyr Phe
GTG GCT GAA GAA GGA AAA GTG TAC GGT TIT CTG AGT TCG AAA ATC GCA CAT CTT
Val Ala Glu Glu Gly Lys Val Tyr Gly Phe Leu Ser Ser Lys Ile Ala His Pro  387 396 405 414 423 432
Phe Phe Ala Val Glu Asp Gly Glu Leu Val Ala Tyr Leu Glu Tyr Phe Asp Val
<b>661</b> 460 450
GAG TTC GAC GAC TTT GTT CCT CTT GAA CCT CTC GTT GTA CTC GAG GAT CCC AAC Glu Phe Anp Asp Phe Val Pro Leu Glu Pro Leu Val Val Leu Glu Asp Pro Asn
ACA CCC CIT CIT CTG GAG AAA TAC GCG GAA CTC GTC GGA ATG GAA AAC AAC GCG
The Pro Leu Leu Glu Lys Tyr Ala Glu Leu Val Gly Met Glu Asn Asn Ala
AGA CTT CUA ANA CAC ACA CCC ACT CGA TCG TCC ACC TCG TAC CAT TAC TTC CTT
Arg Val Pro Lys His The Pro The Gly Trp Cys Ser Trp Tyr His Tyr Phe Leu

Figure 10a.

## Thermotoga maritima Alpha-galactosidade Complete Gene Sequence (2 of 1)

981 990 999 1008 1017 1026 ACC TAC TTC AAG ATC GAC TTT CTC TTC GCG GGT GCC GTT CCA GGA GAA AGA AAA ATG TYT Phe Lys Ile Asp Phe Leu Phe Ala Gly Ala Val Pro Gly Glu Arg Lys UAG AAC ATA ACA CCA ATT CAG GCG TTC AGA AAA GGG ATT GAG ACG ATC AGA AAA AND ASA Ile Thr Pro Ile Gln Ala Phe Arg Lys Gly Ile Glu Thr Ile Arg Lys CC GTG GGA GAA GAT TCT TTC ATC CTC GCA TCC GCC TCT CCC CTT CTT CCC GCA La Val Gly Glu Asp Ser Phe Ile Leu Gly Cys Gly Ser Pro Leu Leu Pro Ala CC GCA TCC GCC GCC ATG AGG ATA GGA CCT CCC CTC TCC GCA CC GCA TCC GCC CCC ATG AGG ATA GGA CCT CAC ACT CCC CTC TCCC GCA CC GCA TCC GCC GCC ATG AGG ATA GGA CCT CAC ACT CCC CTC TCCC GCA CC GCA TCC GCC GCC ATG AGG ATA GGA CCT CAC ACT CCC CCG TTC TCC GCA CC GCA TCC GCC GCC ATG AGG ATA GGA CCT CAC ACT CCC CCG TTC TCC GCA	(
TTC GAG GTC TTC CAG ATA GAC GAC GCC TAC GAA AAG CAC ATA GGT GAC TGC CTC  Phe Glu Val Phe Gln Ile Asp Asp Ala Tyr Glu Lys Asp Ile Gly Asp Trp Leu  711 729 7738 7747 7756  GTG ACA AGA GGA GAC TTT CCA TCC GTG GAA GAG ATG GCA AAA GTT ATA CCG GAA  Val Thr Ary Gly Asp Phe Pro Ser Val Glu Glu Met Ala Lys Val Ile Ala Glu  Val Thr Ary Gly Asp Phe Pro Ser Val Glu Glu Met Ala Lys Val Ile Ala Glu  AMAC GGT TTC ATC CCG GGC ATA TGG ACC GCC CCG TTC AGT GTT TCT GAA ACC TCC  ASTA Gly Phe Ile Pro Gly Ile Trp Thr Ala Pro Phe Ser Val Ser Glu Thr Ser  GAT GTA TTC AAC GAA CAT CCG CAC TGG GTA GTG AAG GAA AAC GGA GAG CCG AAG  Asp Val Phe Asm Glu His Pro Asp Trp Val Val Lys Glu Asm Gly Glu Pro Lys  ATG GCT TAC AGA AAC TGG AAC AAA AAG ATA TAC GCC CTC GAT CTT TCG AAA GAT  ANG GCT TAC AGA AAC TGG AAC AAA AAG ATA TAC GCC CTC GAT CTT TCG AAA GAT  ANG GCT TAC AGA AAC TGG AAC AAA AAG ATA TAC GCC CTC GAT CTT TCG AAA GAT  Met Ala Tyr Ary Asm Trp Asm Lys Lys Ile Tyr Ala Leu Asp Leu Ser Lys Asp  CAG GTT CTG AAC TGG CTT TTC GAT CTC TTC TCA TCT CTG AGA AAG ATG GGC TAC  Glu Val Leu Asm Trp Leu Phe Asp Leu Phe Ser Ser Leu Ary Lys Het Gly Tyr  ACG TTC TA AAG ATC GAC TTT CTC TTC CCG GGT CCC GTT CCA GGA CAA AAG AAA  ATG Tyr Phe Lys Ile Asp Phe Leu Phe Ala Gly Ala Val Pro Gly Glu Arg Lys  AAG TTC AAAA ACA CCA ATT CAG GCC TTC AGA AAA GGA ATA  AATY Tyr Phe Lys Ile Asp Phe Leu Phe Ala Gly Ala Val Pro Gly Glu Arg Lys  AAG AAC ATA ACA CCA ATT CAG GCC TTC AGA AAA GGG ATT GAG ACG ATC AGA AAA  ATG Tyr Phe Lys Ile Asp Phe Leu Phe Arg Lys Gly Ile Glu Thr Ile Arg Lys  AAG AAC ATA ACA CCA ATT CTG GCC TTC AGA AAA GGG ATT GAG ACG ATC AGA AAA  ATG Tyr Phe Lys Ile Asp Phe Leu Gly Cys Gly Ser Pro Leu Leu Pro Ala  CC GTC GCA GAA GAT TCT TTC ATC CTC GCA TCC GCC TTC CCC CTT CTC CCC  ATG GTC GCA GAA GAT TCT TTC ATC CTC GCA TCC GCC TTC CCC CTT CTC CCC  ATG GCA TCC GCC TTC GCA TCC GCC TTC CCC CTT CTC CCC  ATG GCA TCC GCC TTC GCA TCC CCC TTC CTC CCC CTC CTC CCC  ATG GCA TCC GCC TTC GCC ATC ACC ATC ACC ATC CCC CTC CTC CCC  ATG GCA TCC GCC TTC GCC ATC ACC ATC ACC ATC CCC CTC TTC	CAT CTC ACC TOO CAN CAG ACT CTC AAG AAC CTC AAG CTC CTC AAC CTC AAG CTC CTC AAC CTC CTC AAC CTC CTC AAC CTC CT
738 747 756  739 738 747 756  740 757 756  741 757 756  742 757 756  743 757 756  744 753 756  745 756  745 756  746 757 756  746 757 756  747 758  748 758 792 801 810  748 758 792  801 810  810 810  810 810  828 837 846  845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 634 845 845 845 845 845 845 845 845 845 84	TTC GAG GTC TTC CAG ATA GAC GAC GCC TAC GAA AAG GAC ATA GCT GAD
AMC GET TIC ATC CCG GEC ATA TEG ACC GCC CCG TIC AGT GIT TOT CAA ACC TCC  ASD Gly Phe Ile Pro Gly Ile Trp Thr Ala Pro Phe Ser Val Ser Glu Thr Ser  819  828  837  846  855  864  ASD GIA TIC AAC GAA CAT CCG CAC TGG GIA GTG AAG GAA AAC GCA GAG CCG AAG  ASD Val Phe Asn Glu His Pro Asp Trp Val Val Lys Glu Asm Gly Glu Pro Lys  873  882  881  882  881  882  881  900  909  918  ATG GCT TAC AGA AAC TGG AAC AAA AAG ATA TAC GCC CTC GAT CTT TGG AAA GAT  Met Ala Tyr Arg Asn Trp Asn Lys Lys Ile Tyr Ala Leu Asp Leu Ser Lys Asp  CAG GTT CTG AAC TGG CTT TTC GAT CTC TTC TCA TCT CTG AGA AAG ATG GGC TAC  Glu Val Leu Asn Trp Leu Phe Asp Leu Phe Ser Ser Leu Arg Lys Het Gly Tyr  ACG TAC TTC AAG ATC GAC TTT CTC TTC GCG GGT GCC GTT CCA GGA GAA AGA AAA  ATG Tyr Phe Lys Ile Asp Phe Leu Phe Ala Gly Ala Val Pro Gly Glu Arg Lys  1035  1044  1053  1062  1071  1080  AGG AGA ATA ACA CCA ATT CAG GCC TTC AGA AAA GGG ATT GAG ACG ATC AGA AAA  AYS ASN Ile Thr Pro Ile Gln Ala Phe Arg Lys Gly Ile Glu Thr Ile Arg Lys  CC GTC GCA GAA GAT TCT TTC ATC CTC GCA TCC CCC TTC TCC CCC ATC  1089  1098  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009  1009	OTG ACA AGA GGA GAC TIT CCA TCG GTG GAA GAG ATG GCA AAA GTT ATA CCG GAA  Val Thr Arg Gly Asp Phe Pro Ser Val Glu Glu Met Ala Lys Val Ile Ala Glu
B28 B37 B46 B37 B46 B35 B46 B37 B46 B37 B46 B37 B46 B37 B46 B35 B64 B37 B47 B48 B48 B37 B48 B38 B38 B39 B39 B30 B31 B31 B31 B31 B31 B32 B31 B31 B31 B32 B31 B31 B30 B31 B31 B31 B32 B31 B31 B31 B30 B31	AMC COT TIC ATC CCG GGC ATA TGG ACC GCC CCG TIC ACT GIT TCT GAA ACC TCC  ASD Gly Phe Ile Pro Gly Ile Trp Thr Ala Pro Phe Ser Val Ser Glu Thr Ser
ATG GCT TAC AGA AAC TGG AAC AAA AAG ATA TAC GCC CTC GAT CTT TGG AAA GAT  Met Ala Tyr Arg Asn Trp Asn Lys Lys Ile Tyr Ala Leu Asp Leu Ser Lys Asp  227  CAG GTT CTG AAC TGG CTT TTC GAT CTC TCA TCT CTG AGA AAG ATG GGC TAC  Glu Val Leu Asn Trp Leu Phe Asp Leu Phe Ser Ser Leu Arg Lys Met Gly Tyr  AGG TAC TTC AAG ATC GAC TTT CTC TTC GCG GGT GCC GTT CCA GGA GAA AGA AAA  Arg Tyr Phe Lys Ile Asp Phe Leu Phe Ala Gly Ala Val Pro Gly Glu Arg Lys  1035  1044  1053  1062  1071  1080  AGG AAC ATA ACA CCA ATT CAG CCC TTC AGA AAA GCG ATT GAG ACG ATC AGA AAA  AYS ASN Ile Thr Pro Ile Gln Ala Phe Arg Lys Gly Ile Glu Thr Ile Arg Lys  CC GTG GGA GAA GAT TCT TTC ATC CTC GGA TGC GGC TCT CCC CTT  1089  1098  1098  1107  1116  1125  1134  1152  1161  1170  1179  1188  TC GCA TGC GTC GAC GCC TTC TTC GGA ATA GGA CTT TCC GGA  ACT TCC TTC TTC GGA CCC TTC TTC GGA  1179  1188	GAT GIA TIC AAC GAA CAT CCG CAC TGG GIA GIG AAG GAA AAC GGA GAG CCG AAG Asp Val Phe Asm Glu His Pro Asp Trp Val Val Lys Glu Asm Gly Glu Pro Lys
GAG GTT CTG AAC TGG CTT TTC GAT CTC TTC TCA TCT CTG AGA AAG ATG GGC TAC  Glu Val Leu Asn Trp Leu Phe Asp Leu Phe Ser Ser Leu Ary Lys Met Gly Tyr  981  990  999  1008  1017  1026  AGG TAC TTC AAG ATC GAC TTT CTC TTC GGG GGT GCC GTT CCA GGA GAA AGA AAA  Ary Tyr Phe Lys Ile Asp Phe Leu Phe Ala Gly Ala Val Pro Gly Glu Ary Lys  1035  1044  1053  1062  1071  1080  1080  AGA AAC ATA ACA CCA ATT CAG GCG TTC AGA AAA GGG ATT GAG ACG ATC AGA AAA  AND ASS ASS Ile Thr Pro Ile Gln Ala Phe Ary Lys Gly Ile Glu Thr Ile Ary Lys  CC GTG GGA GAA GAT TCT TTC ATC CTC GGA TGC GGC TCT CCC CTT CTT CCC GCA  10 Val Gly Glu Asp Ser Phe Ile Leu Gly Cys Gly Ser Pro Leu Leu Pro Ala  1143  1152  1161  1170  1179  1188  TC GCA TGC GTC GAC GGG ATG AGG ATA AGG ATA GGA CCT CAC ACT CCG CTG TTC TGG GTA	ATG GCT TAC AGA AAC TGG AAC AAA AAG ATA TAC GCC CTC GAT CTT TGG AAA GAT  Met Ala Tyr Arg Asn Trp Asn Lys Lys Ile Tyr Ala Leu Asp Leu Ser Lys Asp
AGG TAC TTC AAG ATC GAC TTT CTC TTC GCG GGT GCC GTT CCA GGA GAA AGA AAA  ATT TYP Phe Lys Ile Asp Phe Leu Phe Ala Gly Ala Val Pro Gly Glu Arg Lys  1035 1044 1053 1062 1071 1080  AGG AAC ATT ACA CCA ATT CAG GCG TTC AGA AAA GGG ATT GAG ACG ATC AGA AAA  Lys Asn Ile Thr Pro Ile Gln Ala Phe Arg Lys Gly Ile Glu Thr Ile Arg Lys  CG GTG GGA GAA GAT TCT TTC ATC CTC GGA TGC GCC TCT CCC CTT CTT CCC GCA  10 Val Gly Glu Asp Ser Phe Ile Leu Gly Cys Gly Ser Pro Leu Leu Pro Ala  1143 1152 1161 1170 1179 1188  TC GCA TGC GTC GAC GCG ATG AGG ATA GGA CCT CAC ACT CCC CTT TTC TGG GGA	CAG GIT CTG AAC TOG CIT TIC GAT CTC TTC TCA TCT CTG AGA AAG ATG GGC TAC  Glu Val Leu Asn Trp Leu Phe Asp Leu Phe Ser Ser Leu Arg Lys Met Gly Tyr
LAG AAC ATA ACA CCA ATT CAG GCG TTC AGA AAA GGG ATT GAG ACG ATC AGA AAA  LYS ASAN Ile Thr Pro Ile Gin Ala Phe Arg Lys Gly Ile Glu Thr Ile Arg Lys  LOSS 1089 1098 1107 1116 1125 1134  CG GTG GGA GAA GAT TCT TTC ATC CTC GGA TGC GCC TCT CCC CTT CTT CCC GCA  La Val Gly Glu Asp Ser Phe Ile Leu Gly Cys Gly Ser Pro Leu Leu Pro Ala  LI43 1152 1161 1170 1179 1188  CG GCA TGC GCC GTC GAC GCG ATG AGG ATA GGA CCT GAC ACT CCG CCG TTC TGG GGA	AGG TAC TTC AAG ATC GAC TTT CTC TTC GGG GGT GCC GTT CCA GGA GAA AGA AAA Arg Tyr Phe Lys Ile Asp Phe Leu Phe Ala Gly Ala Val Pro Gly Glu Arg Lys
TO GIG GGA GAA GAT TOT TTC ATC CTC GGA TGC GGC TCT CCC CTT CTT CCC GCA  La Val Gly Glu Amp Ser Phe 11e Leu Gly Cym Gly Ser Pro Leu Leu Pro Ala  1143 1152 1161 1170 1179 1188  TC GCA TGC GTC GAC GGG ATG AGG ATA AGGA CCT GAC ACT GGG CGG TTC TGG GGA	ANG ANG ATA ACA CCA ATT CAG GCG TTC AGA ANA GGG ATT GAG ACG ATC AGA ANA  LYS ASN Ile Thr Pro Ile Gln Ala Phe Arg Lys Gly Ile Glu Thr Ile Arg Lys
TO GOA TOO CITY GAC COO ATG AGG ATA OGA CCT CAC ACT COO COO TTO TOO GOA	LG STG GGA GAA GAT TCT TTC ATC CTC GGA TGC GGC TCT CCC CTT CTT CCC GCA  La Val Gly Glu Asp Ser Phe Ile Leu Gly Cys Gly Ser Pro Leu Leu Pro Ala
	1143

Figure 10 (Continued)

## Thermotoga maritima Alpha-qalactusidane Complete Gene Sequence (3.443)

1197 1206 1215 1224 1233 1242 GAA CAT ATA GAA GAC AAC CCA GCT CCC GCT GCA ACA TGG GCG CTG AGA AAC GCC
Glu His Ile Glu Asp Asn Cly Ala Pro Ala Ala Arg Trp Ala Leu Arg Asn Ala
1251 1760
ATR ACG AGG TAC TTC ATG CAC GAC AGG TTC TGG CTG AAC GAC CCC GAC TGT CTG  The Thr Arg Tyr Pho Het His Asp Arg Phe Trp Leu Asn Asp Pro Asp Cys Leu
1305 1314
ATA CTG AGA GAG GAG AAA ACG GAT CTC ACA CAG AAG GAA AAG GAG CTC TAC TCG
The Leu Arg Glu Glu Lys Thr Asp Leu Thr Gln Lys Glu Lys Glu Leu Tyr Ser  1359 1368 1377 1386 1395 1404
THE
Tyr Thr Cys Cly Val Leu Asp Asn Met Ile Ile Glu Ser Asp Asp Leu Ser Leu  1413 1422 1431 1440
THE ACK GAT CAT GGA AAA AAG GTT CTG AAA GAA AGG CTC GGA CTC GTG GTT GTT
Val Arg Asp His Gly Lys Lys Val Leu Lys Glu Thr Leu Glu Leu Leu Gly Gly 1467 1476
AGA CCA CGG GTT CAA AAC ATC ATG TCG GAG GAT CTG AGA TAC GAG ATC GTG TCG
The Ary Val Gln Asn Ile Met Ser Glu Asp Leu Ary Tyr Glu Ile Val Ser
TCT GGC ACT CTC TCA CCA AAC GTC AAG ATC GTG GAT CTG AAC ACC ACC ACC ACC ACC ACC ACC ACC AC
one only the Leu Ser Gly Asn Val Lys Ile Val Val Rep Top New Carrier Glu
TAC CAC CTG GAA AAA GAA GGA AAG TCC TCC CTG AAA AAA AGA GTC GTC AAA
are Led Giu Lys Glu Gly Lys Ser Ser Leu Lys Lys Arg Val Val Lys Arg
TAR CAC GCA AGA AAC TTC TAC TTC TAC GAA CAG CCT GAG AGA GAA TGA 3.
Ilu Asp Gly Arg Asn Phe Tyr Phe Tyr Glu Glu Gly Glu Arg Gly

Figure 10c(Continued)

## Thermotoga maritima $\beta$ -mannanase (Sept.)

		•																
			9			18									45			54
5,	ATG	GGG	ATT	GGT	GGC	CYC	GAC	TCC	TGG	AGC	CCG	TCA	GTA	TCG	CCG	GAA	TTC	CII
							<b></b> .											
	Met	Gly	Ile	Gly	Gly	γsÞ	yab	Ser	<b>GIL</b>	5er	Pro	Ser	Val	Ser	λla	Glu	Phe	Leu
			63			72			03			20						
	TTP.	TTTC		للحلات	GAG		<b>T</b> CT		81	~~	بلحلحك	90 GCA	3 CTP	CNC	99			10B
													AG1		GAG	710	GIG	AAA
	Leu	Leu	Ile	Val	Glu	Leu	Ser	Phe	Val	Leu	Phe	Ala	Ser	<b>As</b> D	Glu	Phe	Val	Tare
																	101	Dys
			117		•	126			135			144			153			162
	GTG	GAA	AAC	GGA	λλλ	TTC	GCT	CTG	AAC	GGA	XXX	GAA	TIC	λGλ	TTC	ATT	GGA	AGC
	Val	Glu	Asn	Gly	Lys	Phe	Ala	Leu	Asn	Gly	Lys	Glu	Phe	Arg	Phe	Ile	GŢĀ	Ser
			171			180			189			198			207			2.6
	AAC	AAC		TAC	ATC		ጥልሮ	110		110	CCA	ATG	בידב	CAC				
	λsn	Asn	Tyr	TYI	Het	His	Tyr	Lys	Ser	Asn	Gly	Met	lle	λsp	Ser	Val	Leu	Glu.
			225		•	234			243			252			261			270
	AGT	GCC	YCY	GAC	λТG	GGT	ATA	λλG	GIC	בזכ	AGA	ATC	TGG	CCT	TTC	CIC	CYC	CCC
		21-	>	\	Mar	Clar	710	7	77- 7	7		Ile						
	261	ALA	ALG	rsp	ner	GLY.	116	Dys	VAL	Deu	ALG	116	. IID	GIY	Pne	Leu	Asp	GIA
			279			288			297			306			315			324-
	GAG	AGT	TAC	TGC	λGλ	GAC	λAG	AAC	ACC	TAC	ATG	CAT	CCI	GAG	ccc	GGT	CIT	TTC
	Glu	Ser	TYI	CA2	yra	ysb	ГЛ2	λsn	Thr	īλī	Met	His	Pro	Glu	Pro	Gly	Val	Phe
			333			242			254			260			2.50			
	GGG	cuc		GAA	CCX	342	TCC	AAC	351	CAG	<b>ACC</b>	360 GGT	مكلمة	CAA	369		CNC	378
	Gly	Val	Pro	Glu	Gly	Ile	Ser	Asn	λla	Gln	Ser	Gly	Pbe	Glu	Arg	Leu	Asp	Tyr
																-		
			387			396			405			414			423			432
	ACA	GTT	GCG	AAA	GCG	***	GAA	CIC	GGT	ATA	AAA	CTT	GTC	ATT	GII	CTT	GIG	AAC
	ጥኮታ	Val	Ala	Lvs	Ala	Tare	63	Len	Gly	Tla	Tare	Leu	V-1	T10	32-3	•		
	- ***	***	71214	٠, ٥	,,,,,	<i>D</i> ₃ 3		Deu	GIJ	116	<b>- 13</b> - 3	Deu	447	TIE	ATT	ren	VAI	ASn
			441			450			459			468			477			486
	AAC	TGG	GAC	GAC	TTC	CCT	GGA	ATG	AAC	CAG	TAC	GTG	λGG	TGG	TIT	GGA	GGA	ACC
	Asn	Trp	yeb	Asp	Phe	CJÀ	Gly	Met	λsn	Gln	Tyr	ام۷	Arg	TIP	Phe	Сĵу	CJA	Thr
			495			504			513			522			E2.			540
	CAT	CAC			TTC		AGA	CAT			ATC			GAG	531		110	540 TAC
	His	His	Asp	Asp	Phe	TYE	λrg	Авр	Glu	Lys	Ile	Lys	Glu	Glu	TVI	Lvs	Lvs	Tyr

Figure 11a

		•			•								·· .					
				oga	BAI	itis	en f	)-ma:	Dab		(334	<b>300</b> ;	(=	onti	nue	a) (	رہی رہے	12
	_	54			55	8		56	7	•	57	6	•	58	ς .			
GT	c rc	CI	L CI	C CI	አ አአ	C CA	T GT	כ אא	T AC	C TA	CAC	G GGJ	ست ۱	ים. יי	יים ביים	C 10	594 G GAA	
																C AG	G GAN	
Va.	ı Se			u Va	l As	n Hi	s Va	1 As	n Th	r Ty	r Th	r Gly	/ Val	l Pro	o Ty	r Ar	g Glu	
		60	3		61	2		62	1		63	0		63				
GAG	G CC	C YC	C AT	CAT	C CC	CIC	G GA	G CT	T GC	A AAG	C CA	A CCC	ccc	دون	r rca	~ \~	648 G GAC	
Cl																o ACC	s GAC	
GI	ı Pr	o In	T 11	e Me	t Al	a Tr	p Gl	u Le	וג ט	a Asr	3 Gli	a Pro	Arg	Cys	5 G1:	Th		
		65	7		66	F.		67	=	•		_						
λAJ	TC	G GG	G AA	CAC	יינים		r car	67! ''''' =	- ~~~		684	! 		693	3		702	
										G AAG	GAC	ATG	AGC	TCC	TAC	TA :	702 እ አአር	
Lys	Se:	r Gl	y As	a Thi	Le	ı Va	l Gli	ı Tri	ya'	l lar	Gly						 : Lys	
										,-	·GIL	. net	. Ser	Ser	137	: Ile	: Lys	
		71.			720	)		729	)		738	).		747	,		25.0	
AG1	CIC	GA'	r cc	C AAC	CAC	CIC	c cro	GCT	. CIC	GCG	GAC	GAA	GGA	TTC	אדר	100	756 AAC	
Ser																		
561	Det	ر حم د	PEC	) ASI	Hle	Lev	l Val	Ala	Val	Gly	ysb	Glu	Gly	Phe	Phe	Ser	λsn	
		76	5		774			727			700							
TAC	CY.	/ cc:	TT	: XXX	רכם	TAC	: cc:	' GGA	GAA	رودد	CAC	TO:		801			810 TGG	
JAI	Glu	. CJ	Phe	Lys	Pro	Tyr	Gly	Gly	Glu	Ala	Glu	TED	λla	77/-	1	Cl.,	Trp	
		819												-3-		GLY	irp	
TCC	CCT				828			837			846			855			864	
				166	AAG	AAG	crc	CTI	ICC	ATA	GAG	ACG	CTG	GAC	TTC	CCC	ACG	
	_				_,,	Dys	Deu	red	ser	116	GIU	Thr	Val	qaK	Phe	Cly	Thr	•
		873			882			891			900			909			030	
TTC	CAC	CIC	TAT	CCG	TCC	CAC	TGG	GGT	CIC	AGT	CCA	GAG	AAC	TAT	CCC	CAG	310	
Phe	Hie	1																
		DEU	TYE	PIO	ser	HIS	Trp	Gly	Val	Ser	Pro	Clu	Asn	Tyr	λla	Gln	Trp	
		927			936			945			054							•
GGA	GCG	AAG	TGG	AΤλ	GAA	GAC	CAC	ATA	AAG	ATC.	324	АХА	~ ~	963			972	
												777	GAG	ATC	GGA	λλA	CCC	
Gly	Ala	Lys	Trp	Ile	Glu	Asp	His	Ile	Lys	Ile	Ala	Lys	Glu	Tle	Clv	Lam		
		981							•			-			Gly	Lys	PIO	
بلملات				~	990			999		1	800		1	017		1	026	
				GAA	TAT	GGA	ATT	CCY	AAG	AGT	GCG	CCA	GII	AAC	AGA	ACG	GCC	
Val	Val	Leu	Glu	Glu	TVT	Glv	T30	D==										
			<del>-</del>		- ] -	3	**6	510	Lys	ser	VIS	Pro	Val	Asn	Arg	Thr	Ala	
		1035		1	044		1	053		1	062		1	071			000	
ATC	TAC	AGA	CTC	TGG	YYC	GAT	CTG	GTC	TAC	GAT	CTC	GGT	GGA	GAT	CCz	CC-C:	DBO	
**6	AYI	Arg	ren	TIP	naA	ysb	Leu	Val	Tyr	ABD	Leu	Gly	Gly	Asp	Glv	Ala	Met	

Figure 11b(Continued)

Thermotoga maritima β-mannanase (mac) (continued) (6 CP2)
1089
TTC TGG ATG CTC GCG GGA ATC GGG GAA GGT TCG GAG ATC GA
Phe Trp Met Leu Ala Clu
1143
TAT CCG GAC TAG COO 1161 1170
TAT CCG GAC TAC GAC GGT TTC AGA ATA GTG AAC GAC GAC AGT CCA GAA GCG GAA  TYT Pro Asp Tyr Asp Gly Pho
TYT Pro Asp Tyr Asp Gly Ph
ASP LYT ASP Gly Phe Arg Ile Val Asp Asp Asp
Tyr Pro Asp Tyr Asp Gly Phe Arg Ile Val Asn Asp Asp Ser Pro Glu Ala Glu
1197 1206 1215 1224
CTG ATA AGA GAA TAC GCG AAG CTG TTC AAC ACA GGT GAA GAC ATA AGA GAA GAC Leu Ile Arg Glu Tyr Ale Louis
Leu Ile Arg Clu Car
Leu Ile Arg Glu Tyr Ala Lys Leu Phe Asn Thr Gly Glu Ass Th
Leu Ile Arg Glu Tyr Ala Lys Leu Phe Asn Thr Gly Glu Asp Ile Arg Glu Asp
ACC TGC TGT TTC 1269 1278
1296
Thr Cys Ser Phe Ile Leu Pro Lys Asp Clas No.
Thr Cys Ser Phe Ile Leu Pro Lys Asp Gly Met Glu Ile Lys Lys Thr Val Glu
1305 1314.
1305 1314 1323 1332 1341 1350 GTG AGG GCT GGT GTT TTC GAC TAC AGC AAC ACG TTT GAA AAG TTG TCT GTC AAA Val Arg Ala Gly Val Di
THE AGE AGE TIT GIA 120 TO
Val Arg Ala Gly Val Phe Asp Tyr Ser Asn Thr Phe Gly
Val Arg Ala Gly Val Phe Asp Tyr Ser Asn Thr Phe Glu Lys Leu Ser Val Lys
1359 1368 1377 1386
Val Glu Asp Leu Val Phe Glu Asn Glu Ile Glu Bir Leu Clu
Stu App Leu Val Phe Glu Asn Glu Ile Glu Big Lou Glu
Val Glu Asp Leu Val Phe Glu Asn Glu Ile Glu His Leu Gly Tyr Gly Ile Tyr
GGC TIT GAT CTC CAC 122 1431 1440 1440
1458
Gly Phe Asp Leu Asp Thr Thr Arg Ile Pro Asp Gly Cly William The Arg Ile Pro Asp Cly Cly William The Company of
- J Ju Mis Glu Mor Bha
1467 1476 1485 1494
GAA GGC CAC TIT CAG GGA AAA ACC COO 1494 1503 1513
GAA GGC CAC TIT CAG GGA AAA ACG GTG AAA GAC TCT ATC AAA GCG AAA GTG GTG Glu Gly His Phe Gln Gly Lys Thr Val Lys
GIV His Phe Gln Gly Lys Thr Val Lys Acc Com
Glu Gly His Phe Gln Gly Lys Thr Val Lys Asp Ser Ile Lys Ala Lys Val Val
1521 1530 1539 1548 1557
THE GLA GAG GAA GTT CAT THE TOTAL TOTAL
Lan Glu Ala Arg Tyr Val Leu Ala Glu Glu Val Arg Pho Con Transcription
ISON Glu Ala Arg Tyr Val Leu Ala Glu Glu Val Asp Phe Ser Ser Pro Glu Glu
1575 1584 1593 1600
AAA AAC TCC mon him
THE GON ALL TOG CAG GCA GAG MINO
al Lys Asn Trp Trp Asn Ser Gly Thr Trp Gln Ala Glu Phe Gly Ser Pro Asp
In Ala Glu Phe Gly Ser Pro Asp

Figure 11C(Continued)

		The	TRO	tog		mar	1:1		β-1	BA D	Dan	100	Œ	E2)	 	cont	(	<b>4</b> \	(6 G	- 0
		16	29				_							·	•			u,	6	r [ 1
AT	T GJ	N T	GG A	AC.	GGT	. CY	C CI	<i>G</i> C	CY J	647 AAT	GC	λ GC	165 A CT	6 CX	.G C7	166 NG AA	S (27)	C 33	167. A CT	4
11	e Gl	u 13	א כנו	Sn.	Gly	G1	u Va	1 G	ly .	αεΛ	G1;	y Al	a Le	u Gl	n Le	u As	 n Va	 1 Lv	s Let	-
		108	33			160	•												1726 A CTC	
Pro	G1	y Ly	rs S	er .	Asp	Tri	G1	u G	lu 1	Val	λrç	/ Va	L Al	a Ar	g Ly	s Ph	 - Gl:	 - Ar	A CTC  J Leu	<b>:</b> ·
TCA	GA	A TG	7 T C	AG 2	ATC	1746 CTC	GA(	S TA	17 C C	755 3AC	ATC	TAC	176 . AT	4 T CC:	A AA	177; C GT	3 C GAC		1782	
ser	Gli	ı cy	3 G]	lu )	Ile	Leu	Gl	אנג ז	T X	qe	Ile	Tyr	Ile	Pro	As:	n Val	Glu	Gly	Leu	
AAG	ငေ <u>ာ</u>						170	ا ليان	GG	77	CIG	λλο	ccc	GGC	TG	1827 GTG	ÀAG	ATA	1836 . GGC	
Lys	Gl	/ Ar	g Le	ע מי	πg	Pro	Tyr	. YI	a V	al	Leu	λsn	Pro	G17	TI	Val	Lys	Ile	Gly	
crc							~~~	G:	٠	AA .	AGT	CCC	GYG	ATO	ATC	: ACT	TTC	ಆರ	1890 GGA	
Leu	Asp	Met 1899	e.K	n Л	sn .	Ala	λεπ	Va.	1 G	lu .	Ser	λla	Glu	Ile	Ile	Thr	Phe	Gly	Gly	
AAA	GAG	TAC	λG	λ λ: :	CA :	TTC	CAT	GI	\ A(	<b>&gt;</b> / /	TT	CYC	TTC	CYC	AGA	1935 ACA	CCG	GGG	1944 GTG	
Lys	Glu	Tyr 1953	, yr.	) <b>)</b>	rg I	ed?	His	Val	. <b>λ</b> :	rg 1	ile,	Glu	Phe	λsp	Arg	Thr	λla	Gly	Val	
AAA 	GAA	CTT				~~~	- G11	GIC	GC	er c	AT	CAT	CLC	AGG	TAC	1989 GAT	GGA	CCC 1	1998 ATT	
Lys	CIU	Leu 2007	His	: I)	Le G	ly	Val	Val	Gl	УЯ	gp	His	Leu	λrg	Tyr	Asp	Gly	 Pro	Ile	
TTC :	ATC	GAT	AAT	GI	- A	GA (	CTT	TAT	AA	Aλ	GA .	ACA	GGA	CCT	ATG	2043 TGA	3 •			
Phe :	Ile	Авр	Asn	Va	1 2	rg ]	Leu	TYI	Ly	8 Y	rg :	Thr	Gly 	Gly	 Met					

Figure 11d (Continued)

### AEPII la β-mannosidase (63GR1)

5' ATG CTA CCA GAA GAG TETS CTA 27 36 45
5 STEEL STATE OF STATE OF THE S
Met Leu Pro Glu Glu Phe Leu Trp Gly Val Gly Gln Ser Gly Phe Gln Phe Gli
ATG CCC CAC AND 72 81 90
ATG GGC GAC AAG CTC AGG AGG CAC ATC GAT CCA AAT ACC GAC TGG TGG AAG TGG
Met Gly Asp Lys Leu Arg Arg His Ile Asp Pro Asn Thr Asp Trp Trp Lys Trp
117 126 135
GTT CGC GAT CCT TTC AAC ATA AAA AAG GAG CTT GTG AGT GGG GAC CTT CCC GAG  Val Arg Asp Pro Pho Acc
Val Arg Asp Pro Phe Asn Ile Lys Lys Glu Leu Val Ser Gly Asp Leu Pro Glu
171 180
GAC GGC ATC AAC AAC TAC GAA CTT TTT GAA AAC GAT CAC AAG CTC GCT AAA GGC ASD Gly Ile Asd Asd CTC CTC GCT AAA GGC
Asp Gly Ile Asn Asn Tyr Glu Leu Phe Glu Asn Asp His Lys Leu Ala Lys Gly
225 234 242
CTT GGA CTC AAC GCA TAC AGG ATT GGA ATA GAG TGG AGC AGA ATC TTT CCC TGG
Leu Gly Leu Asn Ala Tyr Arg Ila Gly
Leu Gly Leu Asn Ala Tyr Arg Ile Gly Ile Glu Trp Ser Arg Ile Phe Pro Trp
CCG ACG TGG ACG GTC GAT ACC GAG GTC GAG TTC GAC ACT TAC GGT TTA GTA AAG
Pro Thr Trp Thr Val Asp The Classes
Pro Thr Trp Thr Val Asp Thr Glu Val Glu Phe Asp Thr Tyr Gly Leu Val Lys
GAC GTT AAG ATA GAC AAG TCC ACC CTT GCT GAA CTC GAC AGG CTG GCC AAC AAG  Asp Val Lys Ile Acc to the control of
Asp Val Lys Ile Asp Lys Sor The
Asp Val Lys Ile Asp Lys Ser Thr Leu Ala Glu Lau Asp Arg Leu Ala Asn Lys
GAG GAG GTA ATG TAC TAC AGG CGC GTT ATT CAG CAT TTG AGG GAG CTC GGC TTC
Glu Glu Val Het TVT TOT ACT THE ACT CAG CAG CTC GGC TTC
Glu Glu Val Het Tyr Tyr Arg Arg Val Ile Gln His Leu Arg Glu Leu Gly Phe
AAG GTC TTC GTT AAC CTC AAC CAC TTC TTC TTC TTC TTC T
AAG GTC TTC GTT AAC CTC AAC CAC TTC ACG CTT CCA ATA TGG CTC CAC GAC CCG
Leu Asn His Phe Thr Leu Pro Ile Trp Leu His Asp Pro
495 504 513 522 531 540
ATA GTG GCA AGG GAG AAG GCC CTC ACA AAC GAC AGA ATC GGC TGG GTC TCC CAG
Ile Val Ala Arg Glu Lys Ala Leu Thr Asn Asp Arg Ile Gly Trp Val Ser Gln

Figure 120

									_										
				Y	D.I	I 1	. <b>a</b> β	- 24	200	idas	• (	6301	81)	(co	nti	nued Deux	)		
				19			551	3		56	,								
	λGG	; AC	λG:	rr c	TT	CA(	ملمك ت ادد	, ,,,,	~ ~ ~ ~	- D	, 		570	5	•	58	5		594
										G IA	r GC	r GCT	r TA(	XX	: GCC	CA	CC	כת	594 C GCJ
	Arq	Th	r V	ıl V	al	Ġ1,	, ph		- <del></del>										
							a Pine	: VI	а су	נגנ פ	. YI	a Ala	נאנו ז	Ile	≥ Xla	A His	Al	a Lei	 1 Gly
				3			612												
	GAC	CT			:AC	AC:	1 TYCE			621			63(	)		639	•		648
								, AG	- AC	C 1-10	: AAC	: CY	cca	ATC	GI)	CIL	. CI	G GA	648 CTC
	Asp	Lei	u Va	1 A	s D	Th.	· •				·								
	•				_,			, 261	1 m	r Pne	AST	ı Glu	Pro	Met	: Val	. Val	. Va	l Glu	Leu
			65				666						-						
	GGC	TAG			~	~	000	. ~~		675	) 		684	١.		693	}		702
							· IAC	10	لنانا	K TTT	ccc	: ccc	GCX	GTC	) ATC	λAC	ccc	GAC	702 GCC
	013				14	PIC	ıyr	Ser	G13	Phe	Pro	Pro	Gly	. Val	Met	λsn	Pro	Glu	Ala
			71																
	GCG	22.0			~		720			729			738			747			756
						AIU	CIC	AAC	)TA	XTX	AAC	CCC	CAC	GCC	TIG	GCA	TAT	, γyd	756 ATG
	*****	Lyz		u A	14	TTO	Leu	Yen	Met	: Ile	yes	Aļa	His	λla	Leu	λla	Tyr	Lys	Met
			76			•												-	
	a ጥ a	220		, ,	~		774			783			792			801			810
	717	AAC	, AG	G 1	rc	GAC	ACC	AAG	AAG	GCC	GAT	CYC	CAT	AGC	AAG	TCC	CCI	GCG	810 GAC
	*10	Lys	AI	g P	20	vzb	Thr	Γλa	Lys	Xla	yzb	Glu	yab	Ser	Lys	Ser	Pro	λla	λsp
			81																
	ملحلت				-	m	828			837			846			855			864
				η Д.	11	TAC	AAC	AAC	ATC	GGT	CIT	CCC	TAC	CCT	AAA	GAC	CCI	λλC	864 Gat
		GLY	11	e 13	Le	INI	ASD	Asn	Ile	GIA	Val	Ala	ĮУI	Pro	Lys	λsp	Pro	λsn	λsp
			87																-
	כככ	2 2 0			~		882		_	891			900			909			918
			GA		. 1	AAA	GCA	GCC	GAA	AAÇ	GAC	AAC	TAC	TTC	CAC	AGC	GGA	CIG	918 TTC
		-ys	<b>7</b> 5	Ve	. 1	rys	ATA	Ala	Glu	Asn	ysb	yzu	Tyr	Phe	His	Ser	Gly	Leu	Phe
			921																
	Jelek	СУТ			~	~~~	936			945			954			963			972
				,	_	CAC	AAG	GGT	AAG	CIC	AAC	ATA	GAG	TTC	GAC	GGC	GAA	AAC	TIT
	• •••	ر د.	VIC	. 11		n13	гÀа	GIA	ГЛЗ	Leu	Asn	Ile	Glu	Phe	Asp	Gly	Glu	λsn	Phe
			981															•	
	GTA					~~~	990			999		-1	800		1	1017		1	026
					A .	CAC	CTA	AAA	ccc	AAT	CAC	TGG	λTλ	GGC	CTC	λλС	TAC	TAC	ACC
		<b>⊸y</b> 3	441	. AI	y.	#1 2	ren	Lys	Gly	Asn	qaA	Trp	Il.	Cly	Leu	Asn	Tyr	Tyr	Thr
			1035																
					_	, _. .	044			1053		1	062		1	.071		1	.080
			411	G1	4 4	AGA	TAT	TCG	GAG	222	AAG	TTC	CCA	B ~~			CTC	ATA	TCC
,		- <b></b>			<u> </u>														

Figure 12b(Continued)

Arg Glu Val Val Arg Tyr Ser Glu Pro Lys Phe Pro Ser Ile Pro Leu Ile Ser

### ABPII la β-mannosidase (63GB1) (continued) 1098 TTC ANG GGC GTT CCC ANC THE GGC THE TCC TGC AGG CCC GGC ACG ACC TCC GCC --- --- --- --- --- --- --- --- --- --- --- --- --- ---Phe Lys Gly Val Pro Asn Tyr Gly Tyr Ser Cys Arg Pro Cly Thr Thr Ser Ala GAT GGC ATG CCC GTC AGC GAT ATC GGC TGG GAA GTC TAT CCC CAG GGA ATC TAC 1152 --- --- --- --- --- --- --- --- --- --- ---Asp Gly Met Pro Val Ser Asp Ile Gly Trp Glu Val Tyr Pro Gln Gly Ile Tyr GAC TCG ATA GTC GAG GCC ACC AAG TAC AGT GTT CCT GTT TAC GTC ACC GAG AAC --- --- --- --- --- --- --- --- --- --- --- --- ---Asp Ser Ile Val Glu Ala Thr Lys Tyr Ser Val Pro Val Tyr Val Thr Glu Asn 1251 1260 GGT GTT GCG GAT TCC GCG GAC ACG CTG AGG CCA TAC TAC ATA GTC AGC CAC GTC --- --- --- --- --- --- --- --- --- --- --- --- --- ---Gly Val Ala Asp Ser Ala Asp Thr Leu Arg Pro Tyr Tyr Ile Val Ser His Val 1314 TCA AAG ATA GAG GAA GCC ATT GAG AAT GGA TAC CCC GTA AAA GGC TAC ATG TAC --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---Ser Lys Ile Glu Glu Ala Ile Glu Asn Gly Tyr Pro Val Lys Gly Tyr Met Tyr 1359 1368 TGG GCG CTT ACG GAT AAC TAC GAG TGG GCC CTC GGC TTC AGC ATG AGG TTT GGT --- --- --- --- --- --- --- --- --- --- --- --- ---Trp Ala Leu Thr Asp Asn Tyr Glu Trp Ala Leu Gly Phe Ser Met Arg Phe Gly 1413 CTC TAC AAG GTC GAC CTC ATC TCC AAG GAG AGG ATC CCG AGG GAG AGA AGC GTT 1422 --- --- --- --- --- --- --- --- --- --- --- --- --- ---Leu Tyr Lys Val Asp Leu Ile Ser Lys Glu Arg Ile Pro Arg Glu Arg Ser Val 1467 1476 GAG ATA TAT CGC AGG ATA GTG CAG TCC AAC GGT GTT CCT AAG GAT ATC AAA GAG Glu Ile Tyr Arg Arg Ile Val Gln Ser Asn Gly Val Pro Lys Asp Ile Lys Glu 1530 GAG TTC CTG AAG GGT GAG GAG AAA TGA 3' --- --- --- --- --- ---Glu Phe Leu Lys Gly Glu Glu Lys ***

Figure 12C(Continued)

#### OCI/4V Endoglucanase (33GP1)

18 27 36 45 55 47 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45 56 45	9 18 27
63 72 81 90 99 108 CTC CTA ATC TCA TCC ACT CAG TGT GGA ANA NAT GAA CCA AAC ANA AGA GTG AAT Leu Leu Ile Ser Ser Thr Gln Cys Gly Lys Asn Glu Pro Asn Lys Arg Val Asn 117 126 135 144 153 AGC ATG GAA CAG TCA GTT GCT GAA AGT GAT AGC AAC TCA GCA TTT GAA TAC AAC Ser Met Glu Gln Ser Val Ala Glu Ser Asp Ser Asn Ser Ala Phe Glu Tyr Asn 171 180 189 198 207 216 ANA ATG GTA GGT ANA GGA GTA AAT ATT GGA AAT GCT TTA GAA GGT CCT TTC GAA Lys Met Val Gly Lys Gly Val Asn Ile Gly Asn Ala Leu Glu Ala Pro Phe Glu 225 234 243 252 261 CGA GCT TGG GGA GTA AGT GAG GAT GAA TAT TTT GAG ATA ATA	ATG GTA GAA AGA CAC TTC AGA TAT GTT CTT ATT TGG AGA TAT TGG
TO CITA ATC TCA TCC ACT CAG TGT GGA ANA ANT GAA CCA ANA ANA AGA GTG AAT  Leu Leu Ile Ser Ser Thr Gln Cys Gly Lys Asn Glu Pro Asn Lys Arg Val Asn  117  126  135  144  153  162  AGC ATG GAA CAG TCA GTT GCT GAA AGT GAT AGC AAC TCA GCA TTT GAA TAC AAC  Ser Met Glu Gln Ser Val Ala Glu Ser Asp Ser Asn Ser Ala Phe Glu Tyr Asn  171  180  189  198  207  216  ANA ATG GTA GGT ANA GGA GTA AAT ATT GGA AAT CCT TTA GAA GGT CCT TTC GAA  Lys Met Val Gly Lys Gly Val Asn Ile Gly Asn Ala Leu Glu Ala Pro Phe Glu  225  234  243  252  261  270  GGA GCT TGG GGA GTA AGA ATT GAG GAT GAA TAT TTT GAG ATA ATA	Met Val Glu Arg Nie Db.
TO CITA ATC TCA TCC ACT CAG TGT GGA ANA ANT GAA CCA ANA ANA AGA GTG AAT  Leu Leu Ile Ser Ser Thr Gln Cys Gly Lys Asn Glu Pro Asn Lys Arg Val Asn  117  126  135  144  153  162  AGC ATG GAA CAG TCA GTT GCT GAA AGT GAT AGC AAC TCA GCA TTT GAA TAC AAC  Ser Met Glu Gln Ser Val Ala Glu Ser Asp Ser Asn Ser Ala Phe Glu Tyr Asn  171  180  189  198  207  216  ANA ATG GTA GGT ANA GGA GTA AAT ATT GGA AAT CCT TTA GAA GGT CCT TTC GAA  Lys Met Val Gly Lys Gly Val Asn Ile Gly Asn Ala Leu Glu Ala Pro Phe Glu  225  234  243  252  261  270  GGA GCT TGG GGA GTA AGA ATT GAG GAT GAA TAT TTT GAG ATA ATA	and fine Arg Tyr Val Leu Ile Cys Thr Leu Phe Leu Val Met
117 126 135 144 153 162 AGC ATG GAA CAG TCA GTT GCT GAA AGT GAT AGC AAC TCA GCA TTT GAA TAC AAC SER MET GIU GIN SER VAI Ala GIU SER ASP SER ASN SER ALA Phe GIU TYY ASN 171 180 189 198 207 216 AAA ATG GTA GGT AAA GGA GTA AAT ATT GGA AAT GCT TTA GAA GCT CCT TTC GAA AAT GGA AGT GGT AAA GGA GTA AAT ATT GGA AAT GCT TTA GAA GCT CCT TTC GAA Lys Met Val Gly Lys Gly Val Asn Ile Gly Asn Ala Leu Glu Ala Pro Phe Glu 225 224 234 243 252 261 270 GGA GCT TGG GGA GTA AGA ATT GAG GAT GAA TAT TTT GAG ATA ATA	
117 126 135 144 153 162 AGC ATG GAA CAG TCA GTT GCT GAA AGT GAT AGC AAC TCA GCA TTT GAA TAC AAC SER MET GIU GIN SER VAI Ala GIU SER ASP SER ASN SER ALA Phe GIU TYY ASN 171 180 189 198 207 216 AAA ATG GTA GGT AAA GGA GTA AAT ATT GGA AAT GCT TTA GAA GCT CCT TTC GAA AAT GGA AGT GGT AAA GGA GTA AAT ATT GGA AAT GCT TTA GAA GCT CCT TTC GAA Lys Met Val Gly Lys Gly Val Asn Ile Gly Asn Ala Leu Glu Ala Pro Phe Glu 225 224 234 243 252 261 270 GGA GCT TGG GGA GTA AGA ATT GAG GAT GAA TAT TTT GAG ATA ATA	108
126 135 144 153 162 AGC ATG GAA CAG TCA GTT GCT GAA AGT GAT AGC AAC TCA GCA TTT GAA TAC AAC Ser Met Glu Gln Ser Val Ala Glu Ser Asp Ser Asn Ser Ala Phe Glu Tyr Asn 171 180 189 198 207 216 AAA ATG GTA GGT AAA GGA GTA AAT ATT GGA AAT CCT TTA GAA GCT CCT TTC GAA Lys Met Val Gly Lys Gly Val Asn Ile Gly Asn Ala Leu Glu Ala Pro Phe Glu Lys Met Val Gly Lys Gly Val Asn Ile Gly Asn Ala Leu Glu Ala Pro Phe Glu GGA GCT TGG GGA GTA AGA ATT GAG GAT GAA TAT TTT GAG ATA ATA	Leu Leu Ile Ser Ser Thr Gln Cve Gly Ive
126 135 144 153 162 AGC ATG GAA CAG TCA GTT GCT GAA AGT GAT AGC AAC TCA GCA TTT GAA TAC AAC Ser Met Glu Gln Ser Val Ala Glu Ser Asp Ser Asn Ser Ala Phe Glu Tyr Asn 171 180 189 198 207 216 AAA ATG GTA GGT AAA GGA GTA AAT ATT GGA AAT CCT TTA GAA GCT CCT TTC GAA Lys Met Val Gly Lys Gly Val Asn Ile Gly Asn Ala Leu Glu Ala Pro Phe Glu Lys Met Val Gly Lys Gly Val Asn Ile Gly Asn Ala Leu Glu Ala Pro Phe Glu GGA GCT TGG GGA GTA AGA ATT GAG GAT GAA TAT TTT GAG ATA ATA	the Gly Lys Asn Glu Pro Asn Lys Arg Val Asn
171  180  189  198  207  216  AAA ATG GTA GGT AAA GGA GTA AAT ATT GGA AAT CCT TTA GAA GCT CCT TTC GAA  Lys Met Val Gly Lys Gly Val Asn Ile Gly Asn Ala Leu Glu Ala Pro Phe Glu  225  234  243  252  261  270  GGA GCT TGG GGA GTA AGA ATT GAG GAT GAA TAT TTT GAG ATA ATA	
171  180  189  198  207  216  AAA ATG GTA GGT AAA GGA GTA AAT ATT GGA AAT CCT TTA GAA GCT CCT TTC GAA  Lys Met Val Gly Lys Gly Val Asn Ile Gly Asn Ala Leu Glu Ala Pro Phe Glu  225  234  243  252  261  270  GGA GCT TGG GGA GTA AGA ATT GAG GAT GAA TAT TTT GAG ATA ATA	THE STEET GAN AGT GAT AGC AAC TCA GCA TTT GAN TAC AAC
AMA ATG GTA GGT AMA GGA GTA AMT ATT GGA AMT GCT TTA GAA GCT CCT TTC GAA  Lys Met Val Gly Lys Gly Val Asn Ile Gly Asn Ala Leu Glu Ala Pro Phe Glu  225 234 243 252 261 270  GGA GCT TGG GGA GTA AGA ATT GAG GAT GAA TAT TTT GAG ATA ATA	Ser Met Glu Gln Ser Val Ala Glu Ser Asp Ser Asp Ser Ala Di
AAA ATG GTA GGT AAA GGA GTA AAT ATT GGA AAT GCT TTA GAA GCT CCT TTC GAA  Lys Het Val Gly Lys Gly Val Asn Ile Gly Asn Ala Leu Glu Ala Pro Phe Glu  225 2261 270  GGA GCT TGG GGA GTA AGA ATT GAG GAT GAA TAT TTT GAG ATA ATA	
225 234 243 252 261 270 GGA GCT TGG GGA GTA AGA ATT GAG GAT GAA TAT TTT GAG ATA ATA	ANN ATG GTA GGT ANN GGA GTA ANT ATT GGA ANT GGT TO 207
234 243 252 261 270  GGA GCT TGG GGA GTA AGA ATT GAG GAT GAA TAT TTT GAG ATA ATA	LYS Mer Val Gly Ivo Co
234 243 252 261 270  GGA GCT TGG GGA GTA AGA ATT GAG GAT GAA TAT TTT GAG ATA ATA	var diy bys Gly Val Asn Ile Gly Asn Ala Leu Glu Ala Pro Phe Glu
279  288  297  306  315  324  GGA TTT GAT TCT GTT AGG ATT CCC ATA AGA TGG TCA GCA CAT ATA TCC GAA AAG  Gly Phe Asp Ser Val Arg Ile Pro Ile Arg Trp Ser Ala His Ile Ser Glu Lys  333  342  351  360  369  378  CCA CCA TAT GAT ATT GAC AGG AAT TTC CTC GAA AGA GTT AAC CAT GTT GTC GAT  Pro Pro Tyr Asp Ile Asp Arg Asn Phe Leu Glu Arg Val Asn His Val Val Asp  AGG GCT CTT GAG AAT AAT TTA ACA GTA ATC AAT ACG CAC CAT TTT GAA GAA  Arg Ala Leu Glu Asn Asn Leu Thr Val Ile Ile Asn Thr His His Phe Glu Glu  441  450  459  468  477  486  CTC TAT CAA GAA CCG GAT AAA TAC GGC GAT GTT TTG GTG GAA ATT TCG AGA CAG  Leu Tyr Gln Glu Pro Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Trp Arg Cln  495  504  513  522  531  540  ATT GCA AAA TTC TTT AAA GAT TAC CCG GAA AAT CTG TTC TTT GAA ATC TAC	443 774
279  288  297  306  315  324  GGA TTT GAT TCT GTT AGG ATT CCC ATA AGA TGG TCA GCA CAT ATA TCC GAA AAG  Gly Phe Asp Ser Val Arg Ile Pro Ile Arg Trp Ser Ala His Ile Ser Glu Lys  333  342  351  360  369  378  CCA CCA TAT GAT ATT GAC AGG AAT TTC CTC GAA AGA GTT AAC CAT GTT GTC GAT  Pro Pro Tyr Asp Ile Asp Arg Asn Phe Leu Glu Arg Val Asn His Val Val Asp  AGG GCT CTT GAG AAT AAT TTA ACA GTA ATC AAT ACG CAC CAT TTT GAA GAA  Arg Ala Leu Glu Asn Asn Leu Thr Val Ile Ile Asn Thr His His Phe Glu Glu  441  450  459  468  477  486  CTC TAT CAA GAA CCG GAT AAA TAC GGC GAT GTT TTG GTG GAA ATT TCG AGA CAG  Leu Tyr Gln Glu Pro Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Trp Arg Cln  495  504  513  522  531  540  ATT GCA AAA TTC TTT AAA GAT TAC CCG GAA AAT CTG TTC TTT GAA ATC TAC	CON GCT TOG GGA GTA AGA ATT GAG GAT GAA TAT TIT GAG ATA ATA AAG AAA AGG
GGA TTT GAT TCT GTT AGG ATT CCC ATA AGA TGG TCA GCA CAT ATA TCC GAA AAG Gly Phe Asp Ser Val Arg Ile Pro Ile Arg Trp Ser Ala His Ile Ser Glu Lys  333 342 351 360 369 378  CCA CCA TAT GAT ATT GAC AGG AAT TTC CTC GAA AGA GTT AAC CAT GTT GTC GAT  Pro Pro Tyr Asp Ile Asp Arg Asn Phe Leu Glu Arg Val Asn His Val Val Asp  AGG GCT CTT GAG AAT AAT TTA ACA GTA ATC ATC AAT ACG CAC CAT TTT GAA GAA  Arg Ala Leu Glu Asn Asn Leu Thr Val Ile Ile Asn Thr His His Phe Glu Glu  441 450 459 468 477 486  CTC TAT CAA GAA CCG GAT AAA TAC GGC GAT GTT TTG GTG GAA ATT TGG AGA CAG  Leu Tyr Gln Glu Pro Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Trp Arg Gln  ATT GCA AAA TTC TTT AAA GAT TAC CCG GAA AAT CTG TTC TTT GAA BAC TTG TTG TTG GAA BAC TTG TTG GAA BAC TTG TTG TTG TTG GAA BAC TTG TTG TTG GAA BAC TTG TTG TTG TTG TTG TTG TTG TTG TTG TT	Gly Ala Trp Gly Val Arg Ile Glu Asp Clu Tom Die Gl
GGA TTT GAT TCT GTT AGG ATT CCC ATA AGA TGG TCA GCA CAT ATA TCC GAA AAG  Gly Phe Asp Ser Val Arg Ile Pro Ile Arg Trp Ser Ala His Ile Ser Glu Lys  333  342  351  360  369  378  CCA CCA TAT GAT ATT GAC AGG AAT TTC CTC GAA AGA GTT AAC CAT GTT GTC GAT  Pro Pro Tyr Asp Ile Asp Arg Asn Phe Leu Glu Arg Val Asn His Val Val Asp  AGG GCT CTT GAG AAT AAT TTA ACA GTA ATC ATC AAT ACG CAC CAT TTT GAA GAA  Arg Ala Leu Glu Asn Asn Leu Thr Val Ile Ile Asn Thr His His Phe Glu Glu  441  450  459  468  477  486  CTC TAT CAA GAA CCG GAT AAA TAC GGC GAT GTT TTG GTG GAA ATT TCG AGA CAG  Leu Tyr Gln Glu Pro Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Trp Arg Gln  ATT GCA AAA TTC TTT AAA GAT TAC CCG GAA AAT CTG TTC TTT GAA BCC TTC TTT TTT GAA BCC TTC TTT TTT GAA BCC TTC TTT TTT TTT TTT TTT TTT TTT TT	
333  342  351  360  369  378  CCA CCA TAT GAT ATT GAC AGG AAT TTC CTC GAA AGA GTT AAC CAT GTT GTC GAT  Pro Pro Tyr Asp Ile Asp Arg Asn Phe Leu Glu Arg Val Asn His Val Val Asp  387  396  405  414  423  432  Arg Ala Leu Glu Asn Asn Leu Thr Val Ile Ile Asn Thr His His Phe Glu Glu  441  450  459  468  477  486  CTC TAT CAA GAA CCG GAT AAA TAC GGC GAT GTT TTG GTG GAA ATT TGG AGA CAG  Leu Tyr Gln Glu Pro Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Trp Arg Gln  495  504  513  522  531  540  ATT GCA AAA TTC TTT AAA GAT TAC CCG GAA AAT CTG TTC TTT GAA ATC TTC TTT TTT GAA ATC TTC TTT TTT TTT TTT TTT TTT TTT T	GGA TTT GAT TCT GTT AGG ATT CCC AT 306 315 324
333  342  351  360  369  378  CCA CCA TAT GAT ATT GAC AGG AAT TTC CTC GAA AGA GTT AAC CAT GTT GTC GAT  Pro Pro Tyr Asp Ile Asp Arg Asn Phe Leu Glu Arg Val Asn His Val Val Asp  387  396  405  414  423  432  Arg Ala Leu Glu Asn Asn Leu Thr Val Ile Ile Asn Thr His His Phe Glu Glu  441  450  459  468  477  486  CTC TAT CAA GAA CCG GAT AAA TAC GGC GAT GTT TTG GTG GAA ATT TGG AGA CAG  Leu Tyr Gln Glu Pro Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Trp Arg Gln  495  504  513  522  531  540  ATT GCA AAA TTC TTT AAA GAT TAC CCG GAA AAT CTG TTC TTT GAA ATC TTC TTT TTT GAA ATC TTC TTT TTT TTT TTT TTT TTT TTT T	GIVE DE COL ATA AGA TGG TCA GCA CAT ATA TCC GAA AAG
CCA CCA TAT GAT ATT GAC AGG AAT TTC CTC GAA AGA GTT AAC CAT GTT GTC GAT  Pro Pro Tyr Asp Ile Asp Arg Asn Phe Leu Glu Arg Val Asn His Val Val Asp  387  AGG GCT CTT GAG AAT AAT TTA ACA GTA ATC ATC AAT ACG CAC CAT TTT GAA GAA  Arg Ala Leu Glu Asn Asn Leu Thr Val Ile Ile Asn Thr His His Phe Glu Glu  441  450  459  468  477  486  CTC TAT CAA GAA CCG GAT AAA TAC GGC GAT GTT TTG GTG GAA ATT TGG AGA CAG  Leu Tyr Gln Glu Pro Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Trp Arg Gln  495  504  513  522  531  540  ATT GCA AAA TTC TTT AAA GAT TAC CCG GAA AAT CTG TTC TTT GAA ATC TTG TAC	ory Phe Asp Ser Val Arg Ile Pro Ile Arg Trp Ser Ala His Ile Ser Clu Inc
Pro Pro Tyr Asp Ile Asp Arg Asn Phe Leu Glu Arg Val Asn His Val Val Asp  378  387  396  405  414  423  432  432  Arg Ala Leu Glu Asn Asn Leu Thr Val Ile Ile Asn Thr His His Phe Glu Glu  441  450  459  459  468  477  486  CTC TAT CAA GAA CCG GAT AAA TAC GGC GAT GTT TTG GTG GAA ATT TGG AGA CAG  Leu Tyr Gln Glu Pro Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Trp Arg Gln  495  504  513  522  531  540  ATT GCA AAA TTC TTT AAA GAT TAC CCG GAA AAT CTG TTC TTT GAA ATC	333 343
387 396 405 414 423 432 AGG GCT CTT GAG AAT AAT TTA ACA GTA ATC ATC AAT ACG CAC CAT TTT GAA GAA Arg Ala Leu Glu Asn Asn Leu Thr Val Ile Ile Asn Thr His His Phe Glu Glu 441 450 459 468 477 486 CTC TAT CAA GAA CCG GAT AAA TAC GGC GAT GTT TTG GTG GAA ATT TGG AGA CAG Leu Tyr Gln Glu Pro Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Trp Arg Gln ATT GCA AAA TTC TTT AAA GAT TAC CCG GAA AAT CTG TTC TTT GAA ATC TTG TAC AGA ATC TTT GAA ATC TTT TTT GAA ATC TTT TTT ATC TTT ATC TTT GAA ATC TTT TTT ATC TTT AT	CCA CCA TAT GAT ATT GAC AGG AAT TTC CTC GAA AGA GTT AAC GAD
AGG GCT CTT GAG AAT AAT TTA ACA GTA ATC ATC AAT ACG CAC CAT TTT GAA GAA ATG Ala Leu Glu Asn Asn Leu Thr Val Ile Ile Asn Thr His His Phe Glu Glu  441 450 459 468 477 486 CTC TAT CAA GAA CCG GAT AAA TAC GGC GAT GTT TTG GTG GAA ATT TGG AGA CAG Leu Tyr Gln Glu Pro Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Trp Arg Gln  495 504 513 522 531 540	Pro Pro Tyr Asp Ile Asp and As
AGG GCT CTT GAG AAT AAT TTA ACA GTA ATC ATC AAT ACG CAC CAT TTT GAA GAA ATG Ala Leu Glu Asn Asn Leu Thr Val Ile Ile Asn Thr His His Phe Glu Glu  441 450 459 468 477 486 CTC TAT CAA GAA CCG GAT AAA TAC GGC GAT GTT TTG GTG GAA ATT TGG AGA CAG Leu Tyr Gln Glu Pro Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Trp Arg Gln  495 504 513 522 531 540	And Ash Phe Leu Glu Arg Val Ash His Val Val Asp
441 450 459 468 477 486  CTC TAT CAA GAA CCG GAT AAA TAC GGC GAT GTT TTG GTG GAA ATT TCG AGA CAG  Leu Tyr Gln Glu Pro Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Trp Arg Gln  495 504 513 522 531 540	387 36c
441 450 459 468 477 486  CTC TAT CAA GAA CCG GAT AAA TAC GGC GAT GTT TTG GTG GAA ATT TCG AGA CAG  Leu Tyr Gln Glu Pro Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Trp Arg Gln  495 504 513 522 531 540	AT ANT TTA ACA GTA ATC ATC AAT ACG CAC CAT TTT GAA GAA
CTC TAT CAA GAA CCG GAT AAA TAC GGC GAT GTT TTG GTG GAA ATT TCG AGA CAG  Leu Tyr Gln Glu Pro Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Trp Arg Gln  495 504 513 522 531 540	Arg Ala Leu Glu Asn Asn Leu Thr Val Ile Ile Asn The William
Leu Tyr Gln Glu Pro Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Trp Arg Gln  495 504 513 522 531 540	· WAI AFA
495 504 513 522 531 540	CTC TAT CAA GAA CCG GAT AAA TAC GGC GAT CTT TTC CTC
ATT GCA AAA TTC TTT AAA GAT TAC CCG GAA AAT CTG TTC TTT GAA ATC TTG TAC AAC	Leu Tyr Cla Cla Pro
ATT GCA AAA TTC TTT AAA GAT TAC CCG GAA AAT CTG TTC TTT GAA ATC TTG TAC AAC	old Fro Asp Lys Tyr Gly Asp Val Leu Val Glu Ile Tro Arg Glo
540	447 604
Ile Ala Lys Phe Phe Lys Asp Tyr Pro Glu Asn Leu Phe Phe Glu Ile Tyr Asn	531 540
Ash Leu Phe Phe Glu Ile Tyr Ash	Ile Ala Lys Phe Phe Lys Asp Tyr Pro Clu
	Ash Leu Phe Phe Glu Ile Tyr Ash

Figure 13a

OC1/4V Endoglucanase (33GP1) (continued)
GAG CCT CAC NO 558 567 576 COntinued)
GAG CCT GCT CAG AAC TTG ACA GCT GAA AAA TGG AAC GCA CTT TAT CCA AAA GTG
Glu Pro Ala Gln Asn Leu Thr Ala Glu Lys Trp Asn Ala Leu Tyr Pro Lys Val
Ash Leu Thr Ala Glu Lys Trp Ash Ala Leu Tor Beat
603 612 621
CTC AAA GTT ATC AGG GAG AGC ANT COL
CTC AAA GTT ATC AGG GAG AGC AAT CCA ACC CGG ATT GTC ATT ATC GAT GCT CCA
Leu Lys Val Ile Arg Glu Ser Asp Dro mb.
Leu Lys Val Ile Arg Glu Ser Asn Pro Thr Arg Ile Val Ile Ile Asp Ala Pro
857 666 675 684
ARE THE GCA CAC TAT AGE GCA GTG AGA AGT CTA AND 593 702
AAC TGG GCA CAC TAT AGC GCA GTG AGA AGT CTA AAA TTA GTC AAC GAC AAA CGC
Asn Trp Ala His Tyr Ser Ala Val Arg Ser Leu Lys Leu Val Asn Asp Lys Arg
711 720
ATC ATT GTT TCC TTC 020 729 738 747
ATC ATT GTT TCC TTC CAT TAC TAC GAA CCT TTC AAA TTC ACA CAT CAG GGT GCC
Ile Ile Val Ser Phe His man man and the Court Court Court GCC
Ile Ile Val Ser Phe His Tyr Tyr Glu Pro Phe Lys Phe Thr His Gln Gly Ala
GAA TGG GTT AAT CCC ATC CCA CCT GTT AGG GTT AAG TGG AAT GGC GAG GAA TGG
CLU COLOR OF THE C
Glu Trp Val Asn Pro Ile Pro Pro Val Arg Val Lys Trp Asn Gly Glu Glu Trp
819 828 827
GAA ATT AAC CAA ATC AGA AGT CAT TTC AAA TAC GTG AGT GAC TGG GCA AAG CAA
Glu Ile Asn Gln Ile Arg Ser His Phe Lys Tyr Val Ser Asp Trp Ala Lys Gln
ser his Phe Lys Tyr Val Ser Asp Trp Ala Lys Cla
AAT AAC GTA CCA ATC TIT CIT GGT GAA TIC GGT GGT GGT 909 918
AAT AAC GTA CCA ATC TTT CTT GGT GAA TTC GGT GCT TAT TCA AAA GCA GAC ATG
Asn Asn Val Pro Ile Phe Leu Gly Glu Phe Gly Ala Tyr Ser Lys Ala Asp Het
GAC TCA AGG GTT AAG TGG ACC GAA AGT GTG AGA AAA ATG GCG GAA GAA TTT GGA
GAA AGT GTG AGA AAA ATG GCG GAA GAA TTT CCA
Asp Ser Arg Val Lys Trp Thr Clu Ser W.
Asp Ser Arg Val Lys Trp Thr Glu Ser Val Arg Lys Met Ala Glu Glu Phe Gly
TITT TCA TAC GCG TAT TGG GAA TITT TGT GCA GGA TITT GGC ATA TAC GAT AGA TGG Phe Ser Tyr Ala Tag GAA TGG
Phe Ser The 11
Phe Ser Tyr Ala Tyr Trp Glu Phe Cys Ala Gly Phe Gly Ile Tyr Asp Arg Trp
TCT CAA AAC TOS ATS 014 1053 1062 1071
TCT CAA AAC TOG ATC GAA CCA TTG GCA ACA GCT GTG GTT GGC ACA GGC AAA GAG
Ser Gln Asn Trp Ile Glu Pro Leu Ala Thr Ala Val Val
The Gly The Gly Type Fly
TAA 3

Figure 136(Continued)

#### Thermotoga maritima Pullulanasa (5023)

				_																			
٠.	<b>.</b>			9				18			2	7 .			3	6				_			
,	ATU	GA	T	TT	AC.	A AA	G G	rs o	GG	ATY	C AT	y C	TG	λGG	ب :	YG AI			4:	> 	_		54 333
																		<b>M</b> G	16	G CA	C C	CA	XXX
	net	AS	рI	en	Th	Ly	s Va	11 (	ily	Ile	e Il	e V	al.	λrσ	1.0	ni ke							Lys
														,		<b>u</b> , ,	, 11 G	ııu	1.11	p G1	пА	la	Lys
	C) C			63			7	72			8	1		•	٩	ο.							
	GAL	GT	GG	CA	Aλ)	/ GA	CAC	<b>G</b> 1	TC	ATA	GA	G λ	TA .	AAA	GÀ	C CC	:	. ~	2:	, 			108 TĢG
	>																	~G	GC	r GA	y C	rg	1.èe
	ASD	Va	1 X	la	Lys	) As	P Ar	g P	he	Ile	Gl	u II	le i	Lvs	A e	n 61	v t						Trp
			_											_, _		P 01	y	уs	A19	1 G1	u Va	11	Trp
			_ 1	17			12	6			13	5			14	4							
	ATA	CT	C	AG	GGA	GI	: GX	λG	λG	ATT	TI	TA	C 0	EAA	AA	- A CC	<b>.</b>	. ~	127				162
																		AC	ACA	TC	r cc	C,	AGA
	Ile	Le	ı G	ln	Gly	(a)	Gl	u G	lu	Ile	Phe	<b>1</b> V	<b>T</b> 6	:lu	Lar	. D~						-	
		:	_												٠, د	• ••	U A.	sp	Inr	Sea	Pr	۰.	λrg
			1	71			18	0			189	•			198								
	ATC	TTC	T	IC	GCA	CAG	CC	A A	3G	TCG	AAC	: AA	GG	76	ATC	Ga		~	207			. :	216
																		·T	117	CIG	; yc	כ ג	AAT
	Ile	Phe	P	be .	Ala	Gln	Ala	יע ב	g	Ser	Asn	Ly	s V	al	Ile	G1:		-			· <u>-</u> -		
												_		_			- ~_	·a	rne	Let	Th	r j	rsu
	CC#.			25			234				243				252				261				_
	CCI.	CIG	G	۱T.	ACG	λλλ	AAC	<b>.</b> A?	/Y	GAA	CIC	TT	C A	AG I	GTT		ים י	~ ·	~2~			. ?	270
	D~0	 V-3																'	CAL	CCA	AA.	A C	SAG
	Pro	VAI	Α.5	. d:	Thr	Lys	Lys	: Ly	78	Glu	Leu	Pho	e L	ys '	Val	Thi	·Va	3					
														•					~Sp	GIA	Ly	s C	ilu
	) T~P	~~~	27	y .			288	ŀ			297			:	306				316			_	
	ATT	-	GI	<b>C</b> 7	ICA.	yGY	GTG	GA	λ	AAG	GCC	GAT	c	נ סם	ACG	GAC	יים י		272			3	24
	T10	D																- ·	3AC	GIG	YC	S A	УC
	Ile .	PIO	٧a	1 5	er	yla.	Val	Gl	u I	) YS	λla	λsp	P 2	ro ?	Thr	Ast	71	 د م	\ <del></del>	**- *		· -	
																		_ ′	Sp	val	Thi	: Х	sn .
	ראר (	2445	33		-		342				351			3	360			7	160				
	rac (			^ ^	rac.	CIC	CIT	TC	T C	λλ	TCC	CTG	: AA	LA G	KA	Gλλ	GA		~~~	303		3	18
•	VZ . 1	/a 1	A =	 																VOV.		ı G	AC
	, ΣΣ. <i>I</i>		<b>AL</b> 1	y 1	16	AT	Leu	Se	r G	ilu .	Ser	Leu	Ly	s G	lu	Glu	λsτ	o L	.e11	) TO	T		
			38:																	ar y	гуч	A.	sp
C	TG 0				<del></del>	. m .	396				405			4	14			4	23			٨.	3 2
-	TG C			, <u>,</u>		71X	GAA	GG.	C T	yc i	W	CCC	CC	Άλ	GA.	CIC	ATO		TG	ATY:	C)	3.	~
ν	al G	111	T1	. T	1.	77 -												_				~	
	al G				16 .	rte	CIU	GT7	T	AI I	ys.	Pro	λl	a A	rg	Val	Ile	. M	et 1	Mar	C) v	71	-
			441																		<b>4</b>	11	. •
С	TG G	AC .	כאכ	, 140.	ac -	- A m	450			. 4	159			4	68			4	77			40	
_	TG G				nc 1	MT.	TAC	GAT	. C	CA C	:AG	CIC	CC	A G	CC .	GTA	TAT	. 71	ביי (	~~»	CAC	**	
L	eu A	5D .	<b>.</b>	. T	 																	~^	_
	eu A			3	, 1	AE .	YY	ASP	G)	īλ C	lu	Leu	Gly	y A	la '	Val	Tyr	S	er 1	Prn -	G) 11	1.34	
			495																		u	_y	3
A	CG A	TA :		). )	:a c	ى ماست :	504	<b>~~</b> ~		. 5	13			52	22			5:	31			54	oʻ
_	CG A				· ·		.00	ייככ	CC	:C G	11	CI	AAC	TC	C (	GTA	AAG	G	rg c	TT	כזיר	ىلىڭ مەر	ç
T	ur I	le I	he	۸r	ים ע	a) -	·			· <b>-</b> -								_					_
	ur I	•			3 V	ar 1	TP.	Ser	PI	0 V	al S	er	Lys	Tr	۱ ه	Val	Lys	٧z	1 1	en i	 []	Ph	_
																	_						_

Figure 14a

Thermotoga maritime but
Thermotoga maritima Pullulanase (6GF3) (continued) 549
AAA AAC GGA GAA GAC ACA GAA CCG TAC CAG GTT GTG AAC ATG
Lys Asn Gly Glu Asp Thr Glu Pro Tyr Gln Val Val Asn Met Glu Tyr Lys Gly
AAC GGG GTC TGG GAA GGG GTT GTT CAA GGG GTG GTG GTG GAA GGG GTT GTT CAA GGG GTG GTG GTG GTG GTG GTG GTG GTG G
AAC GGG GTC TGC GAA GCG GTT GTT GAA GGC GAT CTC GAC GGA GTG TTC TAC CTC  Asn Gly Val Trp Glu Ala Val Val Glu Gly Asp Leu Asp Gly Val Phe Tyr Leu  657
657 666 675 684 693
TAT CAG CTG GAA AAC TAC GGA AAG ATC AGA ACA ACC GTC GAT CCT TAT TCG AAA  Tyr Gln Leu Glu Asn Tyr Gly Lyr Tla
Tyr Gln Leu Glu Asn Tyr Gly Lys Ile Arg Thr Thr Val Asp Pro Tyr Ser Lys  711 720 729 738
CCG GTT TAC GCA AAC CAA GAG AGC GCC GTT GTG AAT CTT GCC AGG ACA AAC Ala Val Tyt Ala Agg Agg CC
Ala Val Tyr Ala Asn Asn Gln Glu Ser Ala Val Val Asn Leu Ala Arg Thr Asn 765 774 783
CCA GAA GGA TGG GAA AAC GAC AGG GGA CCG AAA ATC GAA GGA TAC GAA GAC GCG Pro Glu Gly TTD Glu A
819
ATA ATC TAT GAA ATA CAC ATA GCG GAC ATC ACA GGA CTC GAA ATA ACA
Ile Ile Tyr Glu Ile His Ile Ala Asp Ile Thr Gly Leu Glu Asn Ser Gly Val
AAA AAC AAA GGC CTC TAT CTC GGG CTC ACC GAA GAA AAC ACC AND 918
Lys Asn Lys Gly Leu Tyr Leu Gly Leu Thr Glu Glu Asn Thr Lys Gly Pro Gly
927 936 945 954 963 972 GGT GTG ACA ACA GGC CTT TCG CAC CTT GTG GAA CTC GGT GTT ACA CAC GTT CAT G1y Val The The Glasses
Gly Val Thr Thr Gly Leu Ser His Leu Val Glu Leu Gly Val Thr His Val His
ATA CTT CCT TTC TTT GAT TTC TAC ACA GGC GAC GAA CTC GAT AAA GAT TTC GAG  Ile Leu Pro Phe Phe Asp Phe Tyr Thr Gly Asp Glu Leu Asp Lys Asp Phe Glu  1035
1035 1044 1053
TAC TAC AAC TGG GGT TAC GAT CCT TAC CTG TTC ATG CTT CCC
ya Tyr Tyr Asn Trp Gly Tyr Asp Pro Tyr Leu Phe Het Val Pro Glu Gly Arg

Figure 14b(Continued)

Thermotoga maritima Pullulanase (6GP3) (continued) 1089 1098 TAC TCA ACC GAT CCC AAA AAC CCA CAC ACG AGA ATC AGA GAA GTC AAA GAA ATG . Tyr Ser Thr Asp Pro Lys Asn Pro His Thr Arg Ile Arg Glu Val Lys Glu Met 1143 1152 CTC AAA GCC CTT CAC AAA CAC GGT ATA GGT GTG ATT ATG GAC ATG GTG TTC CCT. Val Lys Ala Leu His Lys His Gly Ile Gly Val Ile Met Asp Met Val Phe Pro 1206 CAC ACC TAC GGT ATA GGC GAA CTC TCT GCG TTC GAT CAG ACG GTG CCG TAC TAC --- --- --- --- --- --- --- --- --- --- --- ---His Thr Tyr Gly Ile Gly Glu Leu Ser Ala Phe Asp Gln Thr Val Pro Tyr Tyr 1251 1260 TTC TAC AGA ATC GAC AAG ACA GGT GCC TAT TTG AAC GAA AGC GGA TGT GGT AAC Phe Tyr Arg Ile Asp Lys Thr Gly Ala Tyr Leu Asn Glu Ser Gly Cys Gly Asn 1305 1314 GTC ATC GCA AGC GAA AGA CCC ATG ATG AGA AAA TTC ATA GTC GAT ACC GTC ACC --- --- --- --- --- --- --- --- --- --- --- --- ---Val Ile Ala Ser Glu Arg Pro Met Met Arg Lys Phe Ile Val Asp Thr Val Thr 136R 1377 TAC TGG GTA AAG GAG TAT CAC ATA GAC GGA TTC AGG TTC GAT CAG ATG GGT CTC --- --- --- --- --- --- --- --- --- --- --- --- ---Tyr Trp Val Lys Glu Tyr His Ile Asp Gly Phe Arg Phe Asp Gln Met Gly Leu 1422 ATC GAC ALL AAG ACA ATG CTC GAA GTC GAA AGA GCT CTT CAT AAA ATC GAT CCA 1431 Ile Asp Lys Lys Thr Het Leu Glu Val Glu Arg Ala Leu His Lys Ile Asp Pro 1476 ACT ATC ATT CTC TAC GGC GAA CCG TGG GGT GGA TGG GGA GCA CCG ATC AGG TTT 1485 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---Thr Ile Ile Leu Tyr Gly Glu Pro Trp Gly Gly Trp Gly Ala Pro Ile Arg Phe 1521 1530 GGA AAG AGC GAT GTC GCC GGC ACA CAC GTG GCA GCT TTC AAC GAT GAG TTC AGA 1539 --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---Gly Lys Ser Asp Val Ala Gly Thr His Val Ala Ala Phe Asn Asp Glu Phe Arg 1584 GAC GCA ATA AGG GGT TCC GTG TTC AAC CCG AGC GTC AAG GGA TTC GTC ATG GGA 1593 1602 --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---Asp Ala Ile Arg Gly Ser Val Phe Asn Pro Ser Val Lys Gly Phe Val Het Gly

Figure 14C(Continued)

Thermotors manial
Thermotoga maritima Pullulanase (60P3) (continued)
1629
1629 1638 1647 1656 1665 1674
GGA TAC GGA AAG GAA ACC AAG ATC AAA AGG GGT GTT GTT GGA AGC ATA AAC TAC
GIV TO COL
Tyr Gly Lys Glu Thr Lys Ile Lys Arg Clus Water
Gly Tyr Gly Lys Glu Thr Lys Ile Lys Arg Gly Val Val Gly Ser Ile Asn Tyr
1683 1692 1701 1710 1719 1719
GAC GGA AAA CTC ATC AAA AGT TTC COO 1710 1719 1720
GAC GGA AAA CTC ATC AAA AGT TTC GCC CTT GAT CCA GAA GAA ACT ATA AAC TAC
Asp Gly Lys Leu Ile Lys Son Di
Asp Gly Lys Leu Ile Lys Ser Phe Ala Leu Asp Pro Glu Glu Thr Ile Asn Tyr
1737 1746
GCA GCG TGT CAC GAC AAC CAC ACA CTG TGG GAC AAG AAC TAC CTT GCC GCC AAA Ala Ala Cyg Rig land tack tack tack tack tack tack tack tack
1782
Ala Ala Cys Ris Lon And
ASP ASR HIS Thr Leu Trp Asp Lys Asp Typ Louis
Ala Ala Cys His Asp Asn His Thr Leu Trp Asp Lys Asn Tyr Leu Ala Ala Lys
GCT GAT ARC 333 336 1809 1818 1007
1836
GCT GAT AAG AAA AAG GAA TGG ACC GAA GAA GAA CTG AAA AAC GCC CAG AAA CTG
Ala Asp Lys Lys Glu Trp Thr Glu Glu Glu Leu Lys Asn Ala Gln Lys Leu
the bed bys Asn Ala Gln Lys Leu
1845 1854 1863 1872 1993
TOO ALL CITY CTC ACTO TOOL CO.
1854 1863 1872 1881 1890 GTT GGT GCG ATA CTT CTC ACT TCT CAA GGT GTT CCT TTC CTC CAC GGA GGG CAG Ala Gly Ala Ile Leu Leu Thr Ser Gln Gly Val Dec
Ala Gly Ala Ile Leu Leu Thr Ser Gln Gly Val Pro Phe Leu His Gly Gly Gln
The Phe Leu His Gly Gly Gln
GAC TTC TGC AGG ACG ACG AAT TTC AAC GAC AAC TTC TGC AGG ACG AAT TTC AAC GAC AAC TTC AAC TTC AAC GAC AAC TTC AA
CAC TTC TGC AGG ACG ACG AAT TTC AAC GAC AAC TCC TAC AAC GCC CCT ATC TCG
Asp Phe Cys Arg Thr Thr Asn Phe Asn Asn Asn Asn
Asp Phe Cys Arg Thr Thr Asn Phe Asn Asp Asn Ser Tyr Asn Ala Pro Ile Ser
1953 1962 1971 1980 1980
ATA AAC GGC TTC GAT TAC GAA AGA ANA CTT CAC TTC 1989 1998
ATA AAC GGC TTC GAT TAC GAA AGA AAA CTT CAG TTC ATA GAC GTG TTC AAT TAC  Ile Asn Gly Phe Asp Tyr Gly Are Level
The Asn Gly Phe Asp Tyr Glu Arg Ive Lou Gland
Ile Asn Gly Phe Asp Tyr Glu Arg Lys Leu Gln Phe Ile Asp Val Phe Asn Tyr
2007 2016 2025
CAC AAG GGT CTC ATA AAA CTC AGA AAA GAA CAC CCT GCT TTC AGG CTG AAA AAC
THE TAKE CALL CALL GOT THE AGG CTG AAA AAC
His Lys Gly Leu Ile Lys Leu Arg Lys Glu His Pro Ala Phe Arg Leu Lys Asn
by by bid his Pro Ala Phe Arg Leu Lys Asn
2061 2070 7070
GCT GAA GAG ATC AAA AAA CAC CTG GAA TTT CTC CCG GGC GGG AGA AGA ATA GTT
THE CTC CCG GGC GGG AGA AGA ATA
Ala Glu Glu Ile Lys Lys His Lon Clu
Ala Glu Glu Ile Lys Lys His Leu Glu Phe Leu Pro Gly Gly Arg Arg Ile Val
2115 2124 2123
2115 2124 2133 2142 2151 2160
THE GLA GGT GAT CCC TGG AAA CAG
la Phe Met Leu Lys Asp His Ala Gly Gly Asp Pro Trp Lys Asp Ile Val Val
ALA GLY GLY ASP PRO TEP LYS ASP TIO VAL
Tie val val

Figure 14d(Continued)

### Thermotoga maritima Pullulanasa (5093) (continue

							,	100461	Buedi	
Ile Tyr Asn	Cly .	Asn Leu	Glu Ly	s Thr	Thr T	r Lys	CTG  Leu	2205 CCA GAA  Pro Glu	GGA A	AA TGG
AAT GTG GTT	GTG ;	2232 AAC AGC  Asn Ser	CAG AA	2241 A GCC  B Ala	GGA AC	2250 A GAA  r Glu	GTG  Val	2259 ATA GAA  Ile Glu	NÓO -	2268
GGA ACA ATA	GYY C	2286 TC CAT	CCG CTI	2295 TCC (	GCG TA	2304 C GTT	CTG	2313 TAC AGA	GAG TY	

Figure 14e(Continued)

Figure 154 Thermotoga maritima MSB8 (Clone # 6GP2) Glycosidase

CTT TTA TTG ATC GTT GAG CTC TCT TTC GTT CTC TTT GCA AGT GAC GAG TTC Leu Leu Leu Ile Val Glu Leu Ser Phe Val Leu Phe Ala Ser Asp Glu Phe

GTG AAA GTG GAA AAC GGA AAA TTC GCT CTG AAC GGA AAA GAA TTC AGA TTC Val Lys Val Glu Asn Gly Lys Phe Ala Leu Asn Gly Lys Glu Phe Arg Phe

ATT GGA AGC AAC TAC TAC ATG CAC TAC AAG AGC AAC GGA ATG ATA GAC Ile Gly Ser Asn Asn Tyr Tyr Met His Tyr Lys Ser Asn Gly Met Ile Asp

AGT GTT CTG GAG AGT GCC AGA GAC ATG GGT ATA AAG GTC CTC AGA ATC TGG Ser Val Leu Glu Ser Ala Arg Asp Met Gly Ile Lys Val Leu Arg Ile Trp

GGT TTC CTC GAC GGG GAG AGT TAC TGC AGA GAC AAG AAC ACC TAC ATG CAT Gly Phe Leu Asp Gly Glu Ser Tyr Cys Arg Asp Lys Asn Thr Tyr Met His

CCT GAG CCC GGT GTT TTC GGG GTG CCA GAA GGA ATA TCG AAC GCC CAG AGC Pro Glu Pro Gly Val Pne Gly Val Pro Glu Gly Ile Ser Asn Ala Gln Ser

-GGT TTC GAA AGA CTC GAC TAC ACA GTT GCG AAA GCG AAA GAA CTC GGT ATA Gly Phe Glu Arg Leu Asp Tyr Thr Val Ala Lys Ala Lys Glu Leu Gly Ile

AAA CTT GTC ATT GTT GTG AAC AAC TGG GAC GAC TTC GGT GGA ATG AAC Lys Leu Val lle Val Leu Val Asn Asn Trp Asp Asp Phe Gly Gly Met Asn

CAG TAC GTG AGG TGG TTT GGA GGA ACC CAT CAC GAC GAT TTC TAC AGA GAT Gln Tyr Val Arg Trp Phe Gly Gly Thr His His Asp Asp Phe Tyr Arg Asp

GAG AAG ATC AAA GAA GAG TAC AAA AAG TAC GTC TCC TTT CTC GTA AAC CAT Glu Lys Ile Lys Glu Glu Tyr Lys Lys Tyr Val Ser Phe Leu Val Asn His

GTC AAT ACC TAC ACG GGA GTT CCT TAC AGG GAA GAG CCC ACC ATC ATG GCC Val Asn Thr Tyr Thr Gly Val Pro Tyr Arg Glu Glu Pro Thr Ile Met Ala

TGG GAG CTT GCA AAC GAA CCG CGC TGT GAG ACG GAC AAA TCG GGG AAC ACG Trp Glu Leu Ala Asn Glu Pro Arg Cys Glu Thr Asp Lys Ser Gly Asn Thr

CTC GTT GAG TGG GTG AAG GAG ATG AGC TCC TAC ATA AAG AGT CTG GAT CCC Leu Val Glu Trp Val Lys Glu Met Ser Ser Tyr Ile Lys Ser Leu Asp Pro

AAC CAC CTC GTG GCT GTG GGG GAC GAA GGA TTC TTC AGC AAC TAC GAA GGA Asn His Leu Val Ala Val Gly Asp Glu Gly Phe Phe Ser Asn Tyr Glu Gly

TTC AAA CCT TAC GGT GGA GAA GCC GAG TGG GCC TAC AAC GGC TGG TCC GGT Phe Lys Pro Tyr Gly Glu Ala Glu Trp Ala Tyr Asn Gly Trp Ser Gly

GTT GAC TGG AAG AAG CTC CTT TCG ATA GAG ACG GTG GAC TTC GGC ACG TTC Val Asp Trp Lys Lys Leu Leu Ser Ile Glu Thr Val Asp Phe Gly Thr Phe

CAC CTC TAT CCG TCC CAC TGG GGT GTC AGT CCA GAG AAC TAT GCC CAG TGG His Leu Tyr Pro Ser His Trp Gly Val Ser Pro Glu Asn Tyr Ala Gln Trp

GGA GCG AAG TGG ATA GAA GAC CAC ATA AAG ATC GCA AAA GAG ATC GGA AAA Gly Ala Lys Trp Ile Glu Asp His Ile Lys Ile Ala Lys Glu Ile Gly Lys

CCC GTT GTT CTG GAA GAA TAT GGA ATT CCA AAG AGT GCG CCA GTT AAC AGA Pro Val Val Leu Glu Glu Tyr Gly Ile Pro Lys Ser Ala Pro Val Asn Arg

ACG GCC ATC TAC AGA CTC TGG AAC GAT CTG GTC TAC GAT CTC GGT GGA GAT Thr Ala Ile Tyr Arg Leu Trp Asn Asp Leu Val Tyr Asp Leu Gly Gly Asp

GGA GCG ATG TTC TGG ATG CTC GCG GGA ATC GGG GAA GGT TCG GAC AGA GAC Gly Ala Met Phe Trp Met Leu Ala Gly Ile Gly Glu Gly Ser Asp Arg Asp

GAG AGA GGG TAC TAT CCG GAC TAC GAC GGT TTC AGA ATA GTG AAC GAC GAC Glu Arg Gly Tyr Tyr Pro Asp Tyr Asp Gly Phe Arg Ile Val Asn Asp Asp

AGT CCA GAA GCG GAA CTG ATA AGA GAA TAC GCG AAG CTG TTC AAC ACA GGT Ser Pro Glu Ala Glu Leu Ile Arg Glu Tyr Ala Lys Leu Phe Asn Thr Gly

GAA GAC ATA AGA GAA GAC ACC TGC TCT TTC ATC CTT CCA AAA GAC GGC ATG Glu Asp Ile Arg Glu Asp Thr Cys Ser Phe Ile Leu Pro Lys Asp Gly Met

GAG ATC AAA AAG ACC GTG GAA GTG AGG GCT GGT GTT TTC GAC TAC AGC AAC

Figure 15b (continued)

Glu Ile Lys Lys Thr Val Glu Val Arg Ala Gly Val Phe Asp Tyr Ser Asn

ACG TTT GAA AAG TTG TCT GTC AAA GTC GAA GAT CTG GTT TTT GAA AAT GAG Thr Phe Glu Lys Leu Ser Val Lys Val Glu Asp Leu Val Phe Glu Asn Glu

ATA GAG CAT CTC GGA TAC GGA ATT TAC GGC TTT GAT CTC GAC ACA ACC CGG Ile Glu His Leu Gly Tyr Gly Ile Tyr Gly Phe Asp Leu Asp Thr Thr Arg

ATC CCG GAT GGA GAA CAT GAA ATG TTC CTT GAA GGC CAC TTT CAG GGA AAA Ile Pro Asp Gly Glu His Glu Met Phe Leu Glu Gly His Phe Gln Gly Lys

ACG GTG AAA GAC TCT ATC AAA GCG AAA GTG GTG AAC GAA GCA CGG TAC GTG Thr Val Lys Asp Ser Ile Lys Ala Lys Val Val Asn Glu Ala Arg Tyr Val

CTC GCA GAG GAA GTT GAT TTT TCC TCT CCA GAA GAG GTG AAA AAC TGG TGG Leu Ala Glu Glu Val Asp Phe Ser Ser Pro Glu Glu Val Lys Asn Trp Trp

AAC AGC GGA ACC TGG CAG GCA GAG TTC GGG TCA CCT GAC ATT GAA TGG AAC Asn Ser Gly Thr Trp Gln Ala Glu Phe Gly Ser Pro Asp Ile Glu Trp Asn

GGT GAG GTG GGA AAT GGA GCA CTG CAG CTG AAC GTG AAA CTG CCC GGA AAG Gly Glu Val Gly Asn Gly Ala Leu Gln Leu Asn Val Lys Leu Pro Gly Lys

AGC GAC TGG GAA GAA GTG AGA GTA GCA AGG AAG TTC GAA AGA CTC TCA GAA Ser Asp Trp Glu Glu Val Arg Val Ala Arg Lys Phe Glu Arg Leu Ser Glu

TGT GAG ATC CTC GAG TAC GAC ATC TAC ATT CCA AAC GTC GAG GGA CTC AAG Cys Glu Ile Leu Glu Tyr Asp Ile Tyr Ile Pro Asn Val Glu Gly Leu Lys

GGA AGG TTG AGG CCG TAC GCG GTT CTG AAC CCC GGC TGG GTG AAG ATA GGC Gly Arg Leu Arg Pro Tyr Ala Val Leu Asn Pro Gly Trp Val Lys Ile Gly

CTC GAC ATG AAC AAC GCG AAC GTG GAA AGT GCG GAG ATC ATC ACT TTC GGC Leu Asp Met Asn Asn Ala Asn Val Glu Ser Ala Glu Ile Ile Thr Phe Gly

GGA AAA GAG TAC AGA AGA TTC CAT GTA AGA ATT GAG TTC GAC AGA ACA GCG Gly Lys Glu Tyr Arg Arg Phe His Val Arg Ile Glu Phe Asp Arg Thr Ala

Figure 15C(continued)

GGG GTG AAA GAA CTT CAC ATA GGA GTT GTC GGT GAT CAT CTG AGG TAC GAT Gly Val Lys Glu Leu His Ile Gly Val Val Gly Asp His Leu Arg Tyr Asp

GGA CCG ATT TTC ATC GAT AAT GTG AGA CTT TAT AAA AGA ACA GGA GGT ATG Gly Pro Ile Phe Ile Asp Asn Val Arg Leu Tyr Lys Arg Thr Gly Gly Met

TGA 1991.

END

Figure 15d(continued)

## Figure No. 16/ Thermotoga maritima MSB8(6gb4)

1 ATG AAA AGA ATC GAC CTG AAT CCT TOO	
1 ATG AAA AGA ATC GAC CTG AAT GGT TTC TGG AGC GTT AGG GAT AAC GAA GGG AGA TTT TCG	
1 Met Lys Arg Ile Asp Leu Asn Gly Phe Trp Ser Val Arg Asp Asn Glu Gly Arg Phe Ser	60
61 TIT Can con account	. 20
OAR GGG ACT GTG CCA CCC CCC	
21 Phe Glu Gly Thr Val Pro Gly Val Val Cla 12	120
21 Phe Glu Gly Thr Val Pro Gly Val Val Gln Ala Asp Leu Val Arg Lys Gly Leu Leu Pro	120
121 CAC CCG TAC COMP	40
121 CAC CCG TAC GTT GGG ATG AAC GAA GAT CTC TTC AAG GAA ATA GAA GAC AGA GAG TGG ATC	
and the tyr val Gly Met Asn Glu Asp Leu Phe Lys Glu Tla GAC AGA GAG TGG ATC	180
41 His Pro Tyr Val Gly Met Asn Glu Asp Leu Phe Lys Glu Ile Glu Asp Arg Glu Trp Ile	60
181 TAC GAG AGG GAG TTC GAG TTC AND GAR TAG	
61 TYT Glu Arg Glu Phe Glu Phe Lys Glu Asp Val Lys Glu Gly Glu Arg Val Asp Leu Val	
The Lys Glu Asp Val Lys Glu Gly Glu Arg Val asp town	240
241 TTT GAG CGG CTT TO THE CASE CGG CGG CTT TO THE CASE CGG CGG CTT TO THE CASE CGG CTT TO THE CASE CGG CTT TO THE CASE CGG CGG CTT TO THE CASE CGG CGG CTT TO THE CASE CGG CGG CGG CGG CGG CGG CGG CGG CGG CG	80
THE GAG GTC GAC ACG CTG TCC CAR COL	•
of Phe Glu Gly Val Asp Thr Leu Ser Asp Val Dur Law Acc GGT GTT TAC CTT GGA AGC ACC	300
81 Phe Glu Gly Val Asp Thr Leu Ser Asp Val Tyr Leu Asn Gly Val Tyr Leu Gly Ser Thr	100
301 GAA GAC ATG TTC ATC CAC TAN GOOD	-
301 GAA GAC ATG TTC ATC GAG TAT CGC TTC GAT GTC ACG AAC GTG TTG AAA GAA AAG AAT CAC	
101 Glu Asp Met Phe Ile Glu Tyr Arg Phe Asp Val Thr Asn Val Leu Lys Glu Lys Asn His	350
361 CCG 32G CCG The Ash His	120
ANG GLG TAC ATA AAA TET CCC AND A CO	
Let Lys Val Tyr Ile Lys Ser Pro Ile Arg Val Pro Inc. The GAG CAG AAC TAC GGG	420
John Leu Glu Gln Asn Tyr Gly	140
421 GTC CTC GGC GGT CCT GAA GAT CCC ATC AGA GGA TAC ATA AGA AAA GCC CAG TAT TCG TAC	• .
141 Val Leu Gly Gly Pro Gly New 2	400
141 Val Leu Gly Gly Pro Glu Asp Pro Ile Arg Gly Tyr Ile Arg Lys Ala Gln Tyr Ser Tyr	480
481 GGA TGG GAG TGG	160
TOO GAC TIGG GGT GCC ACA AME	*
GIY ITP ASP Trp Gly Ala Arg Ile Val Thr Ser Gly Ile Tre	540
161 Gly Trp Asp Trp Gly Ala Arg Ile Val Thr Ser Gly Ile Trp Lys Pro Val Tyr Leu Glu	180
541 GTG TAC AGG GCA CGT CTT CAG GAT TCA ACG GCT TAT CTG TTG GAA CTT GAG GGG AAA GAT 181 Val Tyr Arg Ala Arg Leu Gln Asp Ser Thr Ala Tyr Lou Ly	
181 Val Tyr Arg Ala Arg Leu Cla No.	
The Leu Glu Leu Glu Clu Tura	600
601 GCC CTT CTT CTT STY BYS ASP	200
THE AGG GTG AAC GCT THE COLOR	
Ala Leu Val Arg Val Asn Gly Phe Val His Gly Cha GO AAT CTC ATT GTG GAA GTT TAT	560
July Ash Leu Tio Value	220
661 GTA AAC GGT GAA AAG ATA GGG GAA	
661 GTA AAC GGT GAA AAG ATA GGG GAG TTT CCT GTT CTT GAA AAG AAC GGA GAA AAG CTC TTC 7	
Lys Asn Gly Clusters	20
721 GAT GGP GTC 2	40
SOA GIG TTC CAC CTG AAA GAT CTC	
241 Asp Clarate Control of the Contr	
721 GAT GGA GTG TTC CAC CTG AAA GAT GTG AAA CTA TGG TAT CCG TGG AAC GTG GGG AAA CCG 7 241 Asp Gly Val Phe His Leu Lys Asp Val Lys Leu Trp Tyr Pro Trp Asn Val Gly Lys Pro 2	80 -

	781	TAC	CTO	TA	C G	AT T	TC (	مسدي	T-T-C																	
	261	Tyr	Leu	TV	r As	en D	ho 1	U- 1	110	. GT	3 TT	G A	LA G	AC :	TTA	AAC	GGA	GAC	AT	C TA	C AC	A G	AA	GAA		
		Tyr		•		JP .		AGI	Phe	Va.	l Le	u Ly	's A	sp 1	Leu	Asn	Gly	Glu	110	e Tv	r Av		1	Class		
																									28	0
	841	AAG Lys	AAA	AT	C GG	T T	TG ;	AGA	AGA	GTO	AG.	A AT	ר פי	ст .	מר	C > C										
-	281	Lys	Lys	Il	e Gl	y L	eu A	۱rg	Arg	Va 1	Ar	7 71	- U		-70	GAG	CCC	GAT	GA.	(GA)	A GG	A A	AA	ACT	90	0
		Lys						-	_			<b>,</b> ••	e ve	1 L	ın.	Glu	Pro	Asp	Glu	Glu	G1 د	y L	уs	Thr	30	0
																										•
	301	TTC Phe.	Tla	Dh	- GA	A A.	IC A	AC	GGT	GAG	AA	A GT	C TI	C G	CT :	AAG	GGT	GCT	AAC	TGG	ב אר	T ~				
		Phe.	116	Pne	: G.T	u I	le A	sn	Gly	Glu	Lys	Va:	l Ph	e A	la 1	Lys	Glv	Ala	Acn	T	, A1	1 C(	ic.	TCA	960	٥
																									320	)
:	961	GAA . Glu .	AAC	ATC	CT	C AC	GT	GG	TTG	AAG	G N C													•		
:	321	Glu .	Asn	Ile	Le	u Th	r T		T.e	Toro	CAU	GAZ	A GA	T T.	AC (	SAA .	AAG	CTC	GTC	AAA	ATO	GC	'A 3	AGG	1020	1
		Glu .					•	- P	e u	rys	GIA	GI	ı As	P T	yr C	Slu 1	Lys	Leu	Val	Lys	Met	: Al	a 3	Aro	340	
10																									340	'
	43	AGT ( Ser )	300	AAT	ATO	3 AA	C A	rg (	CTC	AGG	GTC	TGG	GGJ	A GO	GA G	GA 2	י ידב	ראר	CZC							
3	41 5	Ser ,	Ala	Asn	Met	: As	n Me	et I	Leu	Arg	Val	Tro	Glv	, G1	י כ	17 2			GAG	AGA	GAG	AT	CI	TC	1080	
												•	,		., .	- Y	T.E.	lyr	GIu	Arg	Glu	Il	e P	he	360	
10	81 7	TAC A	GA	CTC	TGT	GA.	מים ד																		•	
3	61 7	TAC A	ira :	Leu	Cve				.10	GGT	ATC	ATG	GTG	TG	G C	AG G	AT 1	TC.	ATG	TAC	GCG	TG:	rc	тт	1140	
		Yr A	_		Cy J	, vol	בט ע	ע נו.	∕ел (	Gly	Ile	Met	Val	Tr	p G	ln A	sp I	he !	Met	Tyr	Ala	C-2-	- T.	 		
																									380	
114	al G	AA T lu T	AT (	CCG	GAT	CAT	CI	T C	CG 7	rgg ·	TTC	AGA	AAA	CT	ר פי	ר פר	<b>N</b> C C								*	
31	31 G	lu T	yr 1	Pro	Asp	His	Le	u P	ro 1	ים בי	Phe	Ara	Live	1.0		-G A	AC G	AA (	GAG	GCA	AGA	AAG	A:	ΓT	1200	
				•						•			<b>-</b> y 3	שב	u Al	La A	sn G	lu (	3lu .	Ala	Arg	Lys	I	Le	400	
120	ı G	TG A	GA A	LAA.	CTC	AC A	ימידי																			
40	1 V	TG A	ra I	vs	T.e.:	707	. 124		AT C	CC :	rcc .	ATT	GTT	CT	C TG	G T	GC G	GA A	AC ;	uc.	GAA	AAC	AA	.c.	1260	
		al A		., .	264	λig	ту	г н:	ıs P	ro s	Ser	Ile	Val	Let	ı Tr	ъ с)	s G	ly A	sin A	Asn (	Glu	Asn	Δe		420	
120																									420	
126	1 T(	G GC	SA T	TC (	GAT	GAA	TGG	GC	SA A	AT A	TG (	GCC	AGA	AAA	. GT	e es	.T. C									
42	1 T1	p Gl	уР	he ;	Asp	Glu	Trp	G1	у А	sn M	let i	Ala	Ara	Laze	. 1/2	1 1-		31 A	TC F	AC (	CTC	GGA	AA	C	1320	
													5	_,_	, ,	I AS	p G.	y I	le A	sn I	Leu	Gly	As	n	440	
132	l Ac	G CI	C T.	AC (	CTC	TTC	СРТ		T ~																	
44	l Ar	G CT	u T	yr I	eu	Phe	7.22	בנ		-T G	AG A	ATT :	TGT	GCC	GA	A GA	A GA	C C	CG T	CC A	CT (	CCC	TA:	r	1380	
		g Le		,		- 11-	Asp	Pn	e Pi	:0 G	lu I	le (	Cys	Ala	Gli	u Gl	u As	p P	ro s	er 1	hr i	220	TV	-	460	
1383																									100	
	- 16	G CC. P Pro	A TO	C A	GT (	CCA	TAC	GG	CGG	T G	AA A	AA C	CG .	AAC	AGC	- CD:		c c:								
461	Tr	p Pr	o Se	r s	er 1	Pro	Tyr	Gl	y G1	у G:	lu L	vs A	da :	Acn	200		~ ~~	رين دا مند	AA G	GA G	AC ;	(GG	CAC	•	1440	
												•	•		361	GI	ту	B G]	u G	ly A	sp A	urg	His	;	480	
1441	GT	TGC	3 TA	CG	rg 1	raa	እርጥ													٠.						
481	Va]	TGC Trp	TV	r V	- 0 . - 1 .	·	NG 1	GG(	TG	G A7	rg A	AC I	AC (	GAA	AAC	TAC	GA	A AA	A G	AC A	cc G	GA .	محد	: -	.500	
		Tr	2	- •	<u> </u>	LIP.	ser	Gl}	Tr	p Me	t A	sn T	yr (	Slu	Asn	Tyr	Gli	ı Ly	s As	יד סו	מר כ	1 1/2	,,,,,			
150-																									500	
1501	LLC	: ATC	: AG	C G	AG T	TT (	GGA	TTT	CAC	GG	T GO	CT C	כם ר	`A T	CCx	C									•	
501	Phe	Ile	Se	r GI	lu p	he (	31 y	Phe	Glr	1 G1	γ Al	a P	U	i.c	De-	GAG	ACC	AT.	A GA	G TI	CI	TT 1	CA	1	560	
		Ile					-				,	- w P.	. О Н	15	PTO	Glu	Thr	11	e Gl	u Ph	e P	he S	er		520	
1561																										
521	Lys	CCC Pro	G1.	י כי	A	GA (	AG .	ATA	TTC	CA.	T CC	C G	C A	TG	CTG	AAG	CAC	<b>A</b> A	C AA	A Ch	.c c.	רים רי	י מי	,	620	•
	• •	Pro		- 01	u A	rg G	ıu.	Ile	Phe	Hi	s Pr	o Va	1 M	et :	Leu	Lys	His	Ası	1 130	יים פ	u.		~~			
										Fig	ure	16b	(00	nt∹		٠,			- <b>-</b> -y	2 61	11 V č	11 6	TI		540	
												0	, -0		uue	u j										

162	l GC	A (	CAG	GA	A AC	A T	יוד	370																
541	r GJ	уС	Sln	Glu	ı Az	g L	eu	Ile	AGG	Phe	: Ile	Phe	GG/ Gly	A AA: / As:	T TT	T GO e Gl	SA AU Y Ly	AG TO	ST A: 'S Ly	VA GI	AT T SP P	TC G	AC ;	168 56
1681 561	. AG	TT	TT	GTG	TA	TC	TG	TCC	CAG	CTC									C GG e G1					74
 1741												-		Olu	, WIS	1 11	e Ly	s Ph	e Gl	y Va	1 G1	u H	is	58
581	Trp	A	rg	Ser	Arc	Ly	1G '	TAC	AAA Lys	ACG Thr	GCC Ala	GGC Gly	GCT Ala	CTC Leu	TTC Phe	TG	G CA	G TT	C AA( ≥ Ası	GA	C AG	C TG	G 1	800
1801	CCG	GI	C:	TTC	AGO	TG	G 1	rcc	GCD .	CTC	~~~	<u>.</u>												600
											-	•		<b>-</b> 75	AL G	PIO	Lys	Ala	Leu	Tyr	Tyz	Ty:	r 6	360 520
621	GCG	AG	A A	IGA	TTC	TT		٠ ٠																20 40
1921	CTG Leu	CT	3 G	TG (	GGT	GAG	3 C	GA T	ירידי כ															ВО
1981	CGA Arg	GAA	G	AA G	GG	AGA	A	AA C	CT N	~~ ~														0
2041	TGT (	GAG	TI	TG	GT ·	TGA		205 <u>:</u> 685												261	vr.a.	Arg	68	0

Figure 16¢(continued)

# Figure No. 17 Bankia gouldi (37gp4)

	1 A	TG I	AAA	AAA	TAA	<b>C</b> T	A ~	א מיד	ma				_														
	1. M	let 1	vs	Tare	N 0 =	• • •		. A	16 7	TT .	AAA	AG	G C	TT A	CG :	ΤΑΊ	CT	A C	TT	TĢ	TTT	TT	A A	TG	CTG	60	_
-			-, -	∠y s	M5II	rei	u Le	eu M	et I	he i	Lys	Ar	g L	eu T	hr 1	Гуг	Let	u Pr	o L	eu	Phe	ī.e	11 M	۰.	CTG Leu		
																											)
6	1 C	TC 1	CA	CTA	AGT	TC	A GI	ra G	CT C	AA 1	CT	CC.	T G1	a c	4 4 4	A A	~».								GAC		
2	1 L	eu S	er :	Leu	Ser	Sez	r Va	l A	la G	ln s	Ser	Pro	o Va	3 6	1		CA	r GG	כ כ	GT '	TTA	CA	A G'	IT	GAC Asp	120	)
				٠.									- •			ys	HIE	s G1	у А:	rg :	Leu	Gl	v.	1	Asp	40	)
12	1 G	GA A	AC (	GC :	מדד	<b>СТ</b> 1		T 0												•							•
4.	1 G	lv A	sn I	GC Arg	710	t a			IG T	CT G	GA	GAJ	TA A	T A	CG A	GC	TTA	GC	T G	GT /	AAC	AGO	: c:	c	ттт	180	
				Arg :	116	neu	ı As	LA C	as	er G	ly	Glı	ıIl	e Tì	ır s	er	Leu	Al	a G	y J	Asn	Ser	Le	711	Dhe		
																										60	
18:	ı T	GG A	GT A	AT (	CT	GGA	GA	CAC	CT	CC G	AT	TTI	TA	T AZ	TG	CA	CDD	<b>~</b>	T (2)	· .							
6:	l Tı	rp s	er A	sn A	\la	Gly	As	p Th	r s	r A	sp	Phe	Tv	r As	n A	 la	G1	mb.	- 11-		AT	TTT	TI	'A	GCA	240	
						•		•			-		•			- 4	GIU	111	va	L A	sp	Phe	Le	u,	Ala	80	
- 241	G.	A A	AC T	GG 2	LAT .	AGC	TC	<b>а</b> Ст	ጉ አባ	rar n	~ .										_						
81	Gl	u A:	sn T	GG A	sn	Ser	20.	- To		- A	LA A	ATA	GC	T AT	GG	GC (	GTA	AA	A GA	A A	ΑT	TGG	GA	T (	GGC	300	
				xp A			36.	. De	u 11	e A	rg :	Ile	Ala	a Me	t G	ly 1	Val	Lys	Gl	u A	sn	Tip	As	p (	Gly	100	
301																											
	60	A A/	T G	GC I ly I	AT I	ATT	GA:	- AG	T CC	:G C	AG (	GAG	CAJ	GA	A GC	T >	AAA	ATI	· AG	ΔА	ממ	CTT	ስ ጥ	r -	7.7	2.00	
101	62	y As	sn G	ly T	yr :	Ile	Asp	Se	r Pr	o G	ln (	31u	Glr	Gl	u Al	a i	ωvs	Ile	Ar	3 T.		U-1	71.		5A1	360	
							•										•			3 0	ys	val	110	9 ,	sp	120	
361	GC.	A GC	TA	IT G le A	CT 2	AAC	GGC	AT	A TA	T GT	מ מי	מידו	እጥጽ	~										•			
121	Al	a Al	a I	le A	la A	\sn	Glv	116	· 15/	- V.	1 7	3.	71.	GA	I TG	GC	AC	ACT	CA	G	AA (	GCA	GAC	3 1	TA	420	
				le A			•		1	- '		16	116	AS	Tr	рн	is	Thr	Hi	G)	lu 1	Ala	Glu	L	eu	140	
421	TAC	- AC	A C	T C																					•	•	
141	Tvi	- Th	~ ^	AT G	466	CT	GTT	GAC	TT	TT	TA	CC	AGA	ATO	GC	A G	AC	CTA	TAC	. G0	SA C	AT	ACT	. с	cc	480	
	-,,	- 111	- A:	p G	LULA	па	Val	Asp	Ph	e Ph	e T	hr	Arg	Met	Al	аА	sp	Leu	Тух	G1	V A	SD	Thr	P	TO.	160	
481	AAT	GT.	A AI	G TX	T G	AA .	ATT	TAT	AA	GA	G C	CT	ATA	TAC	ר אי	۰ ۸		T~~				_			-		
161	Asn	Va:	l Me	t Ty	r G	lu	Ile	Tyr	Ası	Gl	u P:	ro	Ile	Tyr	G).		o.	166	-	GI	T A	TT	AAG	A	AT	540	
												-		-,-	911	1 31	er :	ırp	Pro	Va	1 1	le	Lys	A	5n	180	
541	TAT	GC	A GA	G CA	A G	TA :	<b>בד</b> ת																				
181	Tyr	Ala	G1	G CA	n V	a] .	716	71-	661	ATZ	A CC	3T	TCT	AAA	GAC	. C	CA (	GAT	AAT	TT	A A	TA .	ATT	G	ΓA	600	
				u Gl	•• ••		116	VIG	GIY	. 114	A A	g :	Ser	Lys	yat	Pı	ro )	Asp	Asn	Le	u I	le :	lle	Va	al	200	
601																											
201	GGI	AC1	AG	C AA	T T	AT 1	rct	CAG	CAA	GTT	GA	T	STA	GCA	TCA	GC	<b>ZA</b> 0	SAC	CCA	ΔT	h T/	~~ /	·				
201	GIY	Thr	Se	r As	n T)	yr S	er	Gln	Gln	Val	As	p (	/al	Ala	Ser	Al	a A	SD	Dro.	71.		(	3MI	A.	. 1.	660	
																		٠٢		111	2 50	er ,	ısp	Th	ır	220	
661	AAT	GTG	GC)	TA:	r ac	тт	TA	CAT	TTT	<b>ጉ</b> ኔፕ		'א כ		-				•									
221	Asn	Val	Ala	Ту	Th	r L	en	Hie	Phe	T	ייט	~ C	CA	TTT	AAC	CC	G C	TA	GAT	AA	T	ra at	GA	AA	T	720	
	Asn			-		-			4 412	ıyr	ΜŢ	et A	та	Phe	Asn	Pr	ОН	lis .	Asp	Ası	Le	u A	rg	As	n	240	
721									•																	•	
	GTA Val	Ale	C1 -	MC/	GC	AT	TA	GAT	AAT	AAT	GT'	T G	CT '	TTG	TTT	GT	T A	CA (	GAA	TGG	; (2/	T 2	<b>~</b> 2	יינע	т.	780	
	Val	- 14 13	GIN	Thr	Al	a L	eu ,	Asp	Asn	Asn	Va:	1 A	la	Leu	Phe	Va:	1 T	hr (	31u	Trn	יטט	· ·	hr	71.	•	260	٠

	781 TTA AAT ACC GO	A CAL CCA COL	
•	261 Leu Asn Thr Gi	GA CAA GGA GAA CCA GAC AAA GAA AGC ACT AAT ACT TGG ATG GC y Gln Gly Glu Pro Asp Lys Glu Ser Thu	C 7000
•		y Gln Gly Glu Pro Asp Lys Glu Ser Thr Asn Thr Trp Met Al	C TIT TTG 840
		TIP MEC AI	a Phe Leu 280
	AAA GAA AAA GG	T ATA AGT CAC GCT AAT TCC TCT	
	81 Lys Glu Lys Gl	T ATA AGT CAC GCT AAT TGG TCT TTG AGT GAC AAA GCT TTT CC y Ile Ser His Ala Asn Trp Ser Leu Ser Asp Lys Ala Phe Pro	T GAA ACA 900
		Ash Trp Ser Leu Ser Asp Lys Ala Phe Pro	Glu Thr 300
	01 GGG TCT GTA CT	T. C	GIU Thr 300
:	Ol Gly Ser value	I CAA GCA GGA CAA GGT GTA TCT GGT TTA ATT AGC AAT AAA CTT	
	and her Ast Ast	l Gln Ala Gly Gln Gly Val Ser Gly Leu Ile Ser Asn Lys Leu	ACA GCC 960
		Leu Ser Asn Lys Leu	Thr Ala 320
9	61 TCT GGT GAA ATT	GTA AAA AAC ATC ATC CAA AAC TGG GAT ACA GAG ACC TCT ACA	
3	21 Ser Gly Glu Ile	Val Lys Asn Ile Ile Gln Asn Trp Asp Thr Glu Thr Ser Thr	GGA CCT 1000
		and the the Gln Asn Trp Asp Thr Glu Thr Ser Thr	GLY Pro
. 10	1 AAA ACA ACA		GIY Pro 340
	1 Lvs Thr Thr G	TGT AGT ACT ATA GAA TGT ATT AGA GCT GCA ATG GAA ACA GCA	
_	bys inr Thr Gln	Cys Ser Thr Ile Glu Cys Ile Arg Dla Na	CAA GCA 1080
		Cys Ser Thr Ile Glu Cys Ile Arg Ala Ala Met Glu Thr Ala	Gln Ala 360
10	1 GGA GAT GAA ATT	ATA ATT CCC com	
3 (	1 Gly Asp Glu Ile	Ile Ile Ala Pro Gly Asn Tyr Asn Phe Gln Asp Lys Ile Gln	GGT GCC 1140
		and Ala Pro Gly Asn Tyr Asn Phe Gln Asp Lys Ile Gln	GGr GCC 1140
114	1 TTT ANC COM		GIY Ala 380
38	Pho has a	GTT TAC CTT TAT GGT AGT GCT AAC GGA AAC AGT ACA AAC CCT :	
	The Ash Arg Ser	Val Tyr Leu Tyr Gly Ser Ala Asn Gly Asn Ser Thr Asn Pro	ATT ATA 1200
	,	oly Ash Ser Thr Ash Pro	lle Ile 400
120	TTA AGA GGC GAA	AGC GCT ACA AAC CCT CCT GTT TTC TCA GGA TTA GAT TAT AAC A	
40	Leu Arg Gly Glu	Ser Ala Thr Asn Pro Pro Val Phe Ser Gly Leu Asp Tyr Asn A	AT GGC 1260
		Ash Pro Pro Val Phe Ser Gly Leu Asp Tyr Ash A	sn Gly 420
126	TAC CTA TTA ACT		420
421	TVT Len Len Den	ATT GAA GGT GAT TAT TGG AAT ATT AAA GAT ATA GAG TTT AAA A	
	-1. Deg Deg Ser 1	the Glu Gly Asp Tyr Trp Asn Ile Lys Asp Ile Glu Phe Lys T	CT GGG 1320
		The Gid Phe Lys T	hr Gly 440
1321	TCT AAA GGT ATT G	TT CTT GAC AAT TCT AAT GGT AGT AAA TTA AAA AAC CTT GTT G	
441	Ser Lys Gly Ile V	al Leu Asp Asp Car has Gi	FT CAT 1380
		al Leu Asp Asn Ser Asn Gly Ser Lys Leu Lys Asn Leu Val Va	al His 460
1381	GAT ATT GGA GAN G		
461	Asp Ile Gly Clu s	AA GCT ATT CAC TTG CGT GAT GGA TCT AGC AAT AAT AGT ATA GA	
	t and only Gid G	lu Ala Ile His Leu Arg Asp Gly Ser Ser Asn Asn Ser Ile As	T GGT 1440
1441		Ash Ser Ile As	p Gly 480
	TGC ACT ATA TAC A	AT ACA GGT AGA ACT AAA CCT GGT TTT GGT GAA GGT TTA TAT GT.	
481	Cys Thr Ile Tyr As	on Thr Gly Arg Thr Luc Des Co	A GGC 1500
		on Thr Gly Arg Thr Lys Pro Gly Phe Gly Glu Gly Leu Tyr Va	l Gly 500
1501	TCA GAT AAA GGD CD	A Chm can	-
501	Ser Asp Lvs Cly Cl	A CAT GAC ACT TAT GAA AGA GCT TGT AAC AAT AAC ACT ATT GAA	
	· -10 GIA GI	n His Asp Thr Tyr Glu Arg Ala Cys Asn Asn Asn Thr Ile Glu	A AAC 1560
1561	TCT has	Ash inr ile Glu	1 Asn 520
	ACC GTT GGA CC	C AAT GTA ACA GCA GAA GGC GTA GAT GTT AAG GAA GGT ACA ATG	
-41	cys Thr Val Gly Pro	o Asn Val Thr Ala Glu Gly Val Asp Val Lys Glu Gly Thr Met	AAC 1620
		old Gly Val Asp Val Lys Glu Gly Thr Met	Asn 540

Figure 17b(continued)

		•
1621	1 ACT ATT ATA AGA AAT TGC GTG TTT TCT GCC CO.	
541	1 ACT ATT ATA AGA AAT TGC GTG TTT TCT GCA GAA GGA ATT TCA GGA GAA AAT AGC TCA GAT 1 Thr 11e 11e Arg Asn Cys Val Phe Ser Ala Glu Gly 11e Ser Gly Glu Asn Ser Ser Asp	1680
	of val fine Ser Ala Glu Gly Ile Ser Gly Glu Asn Ser Ser han	
		560
1681	THE CALL LIA AAA GGD GCC TATE TOTAL	•
561	Ala Phe Ile Asp Leu Lys Gly Ala Tyr Gly Phe Val Tyr Arg Asn Thr Phe Asn Val Asp	1740
-	And Tyr Gly Phe Val Tyr Arg Asn Thr Phe Asn Val Asn	580
. 1741		360
	THE GIA AIA AAT ACT GGA GTA CAG	
581	Gly Ser Glu Val Ile Asn Thr Gly Val Asn Phe Low Ash GGT ACA GGA TTT AAT ACA	1800
	Gly Ser Glu Val Ile Asn Thr Gly Val Asp Phe Leu Asp Arg Gly Thr Gly Phe Asn Thr	600 -
1801		
601	GGT TTT AGA AAT GCA ATA TTT GAA AAT ACA TAT AAC CTT GGC AGT AGA GCT TCA GAA ATT	
	Gly Phe Arg Asn Ala Ile Phe Glu Asn Thr Tyr Asn Leu Gly Ser Arg Ala Ser Glu Ile	1860
		620
1861	TCA ACT GCT CGT AAA AAA CAA GGT TCT CCT GAA CAA ACT CAC GTT TGG GAT AAT ATT AGA	
621	Ser Thr Ala Arg Lys Lys Cla Car Car GAA CAA ACT CAC GTT TGG GAT AAT ATT AGA	1920
	Ser Thr Ala Arg Lys Lys Gln Gly Ser Pro Glu Gln Thr His Val Trp Asp Asn Ile Arg	
		640
1921	AAC CCT AAT TCT GTT GAT TTT CCA ATA AGT GAT GGT ACA GAA AAT CTA GTA AAT AAA TTC	
641	Ash Pro Ash Ser Val Asp Phe Pro Ile Ser Ash Clumber Co	1980
	Asn Pro Asn Ser Val Asp Phe Pro Ile Ser Asp Gly Thr Glu Asn Leu Val Asn Lys Phe	660
661	TGC CCA GAT TGG AAT ATA GAA CCA TGT AAT CCT GTA GAC GAA ACC AAC CAA GCA CCT ACA	2040
	of val Asp Glu Thr hen Cla na -	2040
		680
2041	ATA AGC TTC CTA TCT CCT GTT AAC AAT ATT ACT TTA GTT GAA GGT TAT AAT TTA CAA GTT 11e Ser Phe Leu Ser Pro Val Asp Asp 11e The Year Con GTT 2	_
681	Ile Ser Phe Leu Ser Pro Val Asn Asn Ile Thr Leu Val Glu Gly Tyr Asn Leu Gln Val	100
	Ash Ash Tie Thr Leu Val Glu Gly Tyr Ash Leu Gln Val	700
701 (	GAA GTT AAT GCT ACT GAT GCA GAT GGA ACT ATT GAT AAT GTA AAA CTT TAT ATA GAT AAC 2	
701 (	Glu Val Asn Ala Thr Asp Ala Asp Gly Thr Ile Asp Asn Val Lys Leu Tyr Ile Asp Asn	160
	The Asp Asn	720
2161 ;	AAT TTA GTT AGG CAA ATA AAT TOT	
721 A	AAT TTA GTT AGG CAA ATA AAT TCT ACT TCA TAT AAA TGG GGC CAT TCT GAT TCT CCA AAT 20	220
	Jet Tyr Dys Trp Gly His Com how a	740
2221 A	ACA GAT GAA CTT AAT GGT CTT ACA GAA GGA ACT TAT ACC TTA AAA GCA ATT GCA ACT GAT 22	
741 T	Thr Asp Glu Leu Asn Gly Leu Thr Glu Gly The Tree The AAA GCA ATT GCA ACT GAT 22	80
•	Thr Asp Glu Leu Asn Gly Leu Thr Glu Gly Thr Tyr Thr Leu Lys Ala Ile Ala Thr Asp 7	60
761 A	NAC GAC GGG GCT TCT ACA GAA ACG CAA TTT ACG TTA ACT GTA ATA ACA GAA CAA AGT CCG 23	40
	The the thr Val Ile Thr Clu Cl	
		80
2341 TO	CT GAG AAT TGT GAC TTT AAT ACA CCT TCT TCA ACT GGT TTA GAA GAT TTT GAC ATT AAA 240	
781 Se	er Glu Asn Cys Asp Phe Asn Thr Pro Car Com Time GAA GAT TTT GAC ATT AAA 24	00 .
	er Glu Asn Cys Asp Phe Asn Thr Pro Ser Ser Thr Gly Leu Glu Asp Phe Asp Ile Lys 80	
	AG TIT TCT AAC GIT TIT GAG TTA GGA TCT GGC GGA CCA TCT TTA AGT AAT TTA AAA ACA $$	
	11A AGT AAT TTA AAA ACA 246	50
	Figure 17C(continued)	
	,,	

	. Ly	s P.	he S	er )	lsn	Val	Phe	: G1	u L	eu G	ly	Ser	G1 3	y Gl	y Pr	o s	er _.	 Leu	Se	- As	n Le	eu L	ys	Thr	82
2461 821	TT	TA	T A	TT A	AT	TGG	AAT	TC	- ~,						A TA				•						252( 84(
2521 841	AAC	GG G1	T G	ra c	CT (	SAT Asp	TAT Tyr	TAT	AT	A AI	AT 1	TTA Jeu	AAA Lys	Pro	A AAJ	A AT	T,	ACC Thr	TTT Phe	CAC Gln	TT	T AA e Ly	LA ,	AAT Asn	2580 860
2581 861 2641															Pro	AS	n P	ne .	Asp	Gly	Asp	Ty	r T	TP	2640 880
881	GTA Val														Dys	Inr	As	in )	Asn	Phe	Thr	Ile	יב ב	YI	2700 900
	TTT Phe										•			- 4.2	THE	PTO	Se	T A	sn (	Sln	Ile	Ser	Ly	s	2760 920
921	ATT . Ile :	ACT Thr	GAT Asp	GAT Asp	TC Se	T A	GT A	TT :	AAT Asn	TTT Phe	Ly:	G C	TT ? eu T	Yr I	CCT Pro ,	AST ASD	CC.	T G	CT 1	TA (	GAC Asp	GAA Glu	AC Th:	T	2820 940
821 ;	ATT T	TT	GTG	AGC	GC	T GZ	AA G	ኳጥ c	: A A											T 28	970 956				

Figure 17d(continued)

## Figure No. 18& Pyrococcus furiosus VCI(7EG1)

lea	leader sequence: amino acids 1-24																				
										2	7		3	6			45				
5 '	ΑT	G A	.GC	AA	S AA	A AA	G TT	C GI	C AT	C GI	A TO	T A	רכ דיז	מ מי	מ בס	TC C		, eec.			54
	Me	t S	er	Lys	s Ly	s Ly	s Ph	e Va	1 11	e Va	1 Se	r I	le Le	u T	hr T	10 C	a	1 13		A C	AG
																- L	eu j	7 <i>6</i>	ı va	ı G	ıın
				63	3		7:	2	•	8	1		•	0							
٠	GC	A A	TA	TAT	TT	T GT	A GA	A AA	G TA			יר זיר	T GA	C C			99			1	8 0
	Αl	a I	le	Туг	Ph	e Va	l Gl	u Ly	s Ty	r Hi	s Th	r Se	r Gl		AC AZ	AG T	LA A	ICI	TC	A A	AT
									•					u A	b r	/5 Se	er 1	.hr	Se	r A	sn
				117	•		126	5		13	_		• •	_							
	AC	CT	CA.	TCT	AC	A CC		_	A AC:	בב ב	) )	T TO	14 C AC	4 		15	3			10	62
	Th	r Se	er	Ser	Thi	Pro	Pro	Gli	n Thi	r Th	- To	1 10	r Th	r Ac	CA	IG G7	T C	:TC	AAC	A.	TT
										- 111.	r ne	u se	r in	r Th	r Ly	's Va	l L	eu	Lys	: I:	le
				171			180	,													
	AG	A TA	4C			' GAC			- TCC	189			191	В		20	7			23	16
	Arg	T)	r	Pro	Ast	Ast	. G) 1	COA	3 IGC	5 CC/	A GG	A GC	T CC	TA 1	T GA	AA T.	G G	ΑT	GGI	G.	AT.
		_					, 013	GI	4 11	PIC	o GI	Y Al	a Pro	o Il	e As	p Ly	s A	sp	Gly	As	q
				225			224														
	GGG	. AA			GNN	- Turn-C	234			243			252	?		26	1			27	0
	Gly	As	n	Pro	Glu	Dhe	TAC	ATI	GAA	ATA	AAC	CT	A TGG	AA	C AT	T CT	T A	ΑT	GCT	AC	T
	•					7110	TYL	TIE	GIU	. Ile	Asn	ı Leı	ı Trp	As:	n Il	e Le	u A:	sn	Ala	Th	r
				279																	
	GGA	4-1-			C > C	200	288			297			306			31	5			32	4
	Glv	Ph	_	Ala	Clu	ATG	ACG	TAC	AAT	TTA	ACC	AGO	GGC	GT	CT.	CA(	C TA	/C	GTC	CA	A
	1		•	n. a	GIU	Mec	Inr	Tyr	Asn	Leu	Thr	Ser	Gly	Va:	l Le	Hi:	Ty	T	Val	Gl	n
																					•
	ת מי			333			342			351			360			369	•			37	8
	G) n	T.O.		JAC Nam	AAC	ATT	GTC	TTG	AGG	GAT	AGA	AGT	AAT	TGG	GTO	CAT	GG	A	TAC	CC	c
	0111	שבו	u 2	Asp	Asn	Ile	Val	Leu	Arg	Asp	Arg	Ser	Asn	TI	Val	His	Gl	Y	Tyr	Pro	5
													•								
,	~~ ~			387			396			405			414			423	1			432	2
	JAA	AT	1 7	TC	TAT	GGA	AAC	AAG	CCA	TGG	AAT	GCA	AAC	TAC	GCA	ACI	' GA	T (	GGC		
(	ilu	116	F	he	Tyr	Gly	Asn	Lys	Pro	Trp	Asn	Ala	Asn	Tyr	Ala	Thr	As	g (	Glv	Pro	
																		•	2		
				41			450			459			468			477				486	
7	ATA	CCA	T	TA (	CCC	AGT	AAA	GTT	TCA	AAC	CTA	ACA	GAC	TTC	Tat			ימ			
Į	le	Pro	L	eu ]	Pro	Ser	Lys	Val	Ser	Asn	Leu	Thr	Asp	Phe	TVY	Len	The	m	KIC.	100	
													•		-3-	₽eu.	111		rie	ser	

TAT AAA CTT GAG CCC AAG AAC GGC CTG CCA ATT AAC TTC GCA ATA GAA TCC TGG
Tyr Lys Leu Glu Pro Lys Asn Gly Leu Pro Ile Asn Phe Ala Ile Glu Ser Trp

TTA ACG AGA GAA GCT TGG AGA ACA ACA GGA ATT AAC AGC GAT GAG CAA GAA GTA
Leu Thr Arg Glu Ala Trp Arg Thr Thr Gly Ile Asn Ser Asp Glu Gln Glu Val

603 612 621 630 639 648

ATG ATA TGG ATT TAC TAT GAC GGA TTA CAA CCG GCT GGC TCC AAA GTT AAG GAG

Met Ile Trp Ile Tyr Tyr Asp Gly Leu Gln Pro Ala Gly Ser Lys Val Lys Glu

657 666 675 684 693 702
ATT GTA GTC CCA ATA ATA GTT AAC GGA ACA CCA GTA AAT GCT ACA TTT GAA GTA
Ile Val Val Pro Ile Ile Val Asn Gly Thr Pro Val Asn Ala Thr Phe Glu Val

711 720 729 738 747 756
TGG AAG GCA AAC ATT GGT TGG GAG TAT GTT GCA TTT AGA ATA AAG ACC CCA ATC
Trp Lys Ala Asn Ile Gly Trp Glu Tyr Val Ala Phe Arg Ile Lys Thr Pro Ile

765 774 783 792 801 810
AAA GAG GGA ACA GTG ACA ATT CCA TAC GGA GCA TTT ATA AGT GTT GCA GCC AAC
Lys Glu Gly Thr Val Thr Ile Pro Tyr Gly Ala Phe Ile Ser Val Ala Ala Asn

819 828 837 846 855 864
ATT TCA AGC TTA CCA AAT TAC ACA GAA CTT TAC TTA GAG GAC GTG GAG ATT GGA
Ile Ser Ser Leu Pro Asn Tyr Thr Glu Leu Tyr Leu Glu Asp Val Glu Ile Gly

ACT GAG TTT GGA ACG CCA AGC ACT ACC TCC GCC CAC CTA GAG TGG TGG ATC ACA
Thr Glu Phe Gly Thr Pro Ser Thr Thr Ser Ala His Leu Glu Trp Trp Ile Thr

927 936 945 954

AAC ATA ACA CTA ACT CCT CTA GAT AGA CCT CTT ATT TCC TAA 3'

Asn Ile Thr Leu Thr Pro Leu Asp Arg Pro Leu Ile Ser *

Figure 18b(continued)

International application No. PCT/US97/22623

· · · · · · · · · · · · · · · · · · ·													
A. CLASSIFICATION OF SUBJECT MATTER  IPC(6) :C07H 21/04; C12N 1/20, 1/14, 5/00, 9/38, 9/42; C08B 30/04  US CL :435/207, 209, 252.3, 254.11, 274, 275, 320.1, 325; 536/23.2  According to International Patent Classification (IPC) or to both national classification and IPC													
B. FIELDS SEARCHED	manonar oragorifoation and II C												
Minimum documentation searched (classification system followed	by classification symbols)												
U.S. : 435/207, 209, 252.3, 254.11, 274, 275, 320.1, 325;													
Documentation searched other than minimum documentation to the e	extent that such documents are included	in the fields searched											
Electronic data base consulted during the international search (nan	me of data base and, where practicable	, search terms used)											
Please See Extra Sheet.													
C. DOCUMENTS CONSIDERED TO BE RELEVANT													
		Dalaman Andria No.											
Category* Citation of document, with indication, where app	propriate, of the relevant passages	Relevant to claim No.											
X GRABNITZ et al. Structure of the β-	-Glucosidase Gene bglA of	1-3, 5											
Clostridium thermocellum: Sequence An		species II											
A of Cellulases and β-Glycosidases Including													
Hydrolase. Eur. J. Biochem. September	er 1991, Vol. 200, No. 2,	4, 6-11											
pages 301-309, see entire document.		:											
X VOORHORST et al. Characterization o	of the celB Gene Coding for	1-3, 5											
β-Glucosidase from the Hyperthermop	_	species I and III											
A furiosus and Its Expression and Site-Dire													
coli. J. Bacteriol. December 1995, Vol	1. 177, No. 24, pages 7105-	4, 6-11											
7111, see entire document.	,												
	·												
		•											
Further documents are listed in the continuation of Box C.	See patent family annex.												
* Special categories of cited documents:  *A* document defining the general state of the art which is not considered	T' later document published after the interdate and not in conflict with the app	ication but cited to understand											
to be of particular relevance	the principle or theory underlying the "X" document of particular relevance; th												
"L" document which may throw doubts on priority claim(s) or which is	considered novel or cannot be consider when the document is taken alone												
cited to establish the publication date of another citation or other	"Y" document of particular relevance; th	e claimed invention cannot be											
"O" document referring to an oral disclosure, use, exhibition or other means	considered to involve an inventive combined with one or more other suc being obvious to a person skilled in	h documents, such combination											
*P* document published prior to the international filing date but later than	"&" document member of the same paten												
Date of the actual completion of the international search	Date of mailing of the international se	arch report											
·	<b>2</b> 1 APR 1998												
26 MARCH 1998	Ø # AFR 1990												
Name and mailing address of the ISA/US	Authorized officer	· h											
		wb											

International application No. PCT/US97/22623

Box I C	Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)
This inter	mational report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:
1	Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
2.	Claims Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3.	Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box II O	Observations where unity of invention is lacking (Continuation of item 2 of first sheet)
	mational Searching Authority found multiple inventions in this international application, as follows:
	ase See Extra Sheet.
•	
1.	As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2.	As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
	As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
1-1	11, species I-III
	·
4.	No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Remark o	n Protest
	No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet(1))(July 1992)*

International application No. PCT/US97/22623

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)
This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:
1. Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
2. Claims Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)
This International Searching Authority found multiple inventions in this international application, as follows:
Please See Extra Sheet.
1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. X As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:  1-11, species I-III
4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Remark on Protest  The additional search fees were accompanied by the applicant's protest.  No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet(1))(July 1992)*

International application No. PCT/US97/22623

### **B. FIELDS SEARCHED**

Electronic data bases consulted (Name of data base and where practicable terms used):

APS and STN (Bioscience and Patent Indexes): Desulfurococc##, Staphylotherm##, Thermatoga, galactosidase#, glucosidase#, beta galactosidase#, beta glucosidase#. Genbank, EMBL, ESTs1-4, STS, N-Geneseq: Seq. ID Nos.: 1-3 and A-Geneseq, PIR, Swissprot: Seq ID Nos.: 15-17.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING This ISA found multiple inventions as follows:

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack Unity of Invention because they are not so linked as to form a single inventive concept under PCT Rule 13.1. The species are as follows: there are 18 distinct enzymes disclosed in the description, as enumerated in Figs. 1-18 and Table 1.

The claims are deemed to correspond to the species listed above in the following manner: while all the claims form one Group for examination, each of the claims is generic to the 18 enzyme species disclosed.

The species listed above do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: each enzyme is a different product, thus has the special technical feature of the recited enzyme, which the other species lack.

Form PCT/ISA/210 (extra sheet)(July 1992)*

International application No. PCT/US97/22623

#### **B. FIELDS SEARCHED**

Electronic data bases consulted (Name of data base and where practicable terms used):

APS and STN (Bioscience and Patent Indexes): Desulfurococc##, Staphylotherm##, Thermatoga, galactosidase#, glucosidase#, beta galactosidase#, beta glucosidase#. Genbank, EMBL, ESTs1-4, STS, N-Geneseq: Seq. ID Nos.: 1-3 and A-Geneseq, PIR, Swissprot: Seq ID Nos.: 15-17.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING This ISA found multiple inventions as follows:

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack Unity of Invention because they are not so linked as to form a single inventive concept under PCT Rule 13.1. The species are as follows: there are 18 distinct enzymes disclosed in the description, as enumerated in Figs. 1-18 and Table 1.

The claims are deemed to correspond to the species listed above in the following manner: while all the claims form one Group for examination, each of the claims is generic to the 18 enzyme species disclosed.

The species listed above do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: each enzyme is a different product, thus has the special technical feature of the recited enzyme, which the other species lack.

Form PCT/ISA/210 (extra sheet)(July 1992)*

	·							•	•
	. •		•			-			• ·
								-	
0									
			·						
					. :				
				·				•	
			*						
					-				
		2						÷	
		·.			-	•			
		•			.*				
	, a								
						÷ *	•		
:				•		,		.*	
	·								
								٠.	